Exam prep

Recitation 10

MIT - 6.802 / 6.874 / 20.390 / 20.490 / HST.506 - Spring 2021

Jackie Valeri

Onto recitation R10!

A. Spring 2017
   - Problem 2a-e
   - Problem 3a,b
   - Problem 4a-e

B. Spring 2019
   - Problem 1a-c

C. Spring 2020
   - Problem 4a-d

Spring 2017: Problem 2

This problem makes use of the STL-10 dataset, which contains 500 training images and 800 test images for each of 10 different classes: airplane, bird, car, cat, deer, dog, horse, monkey, ship, and truck. You decide to construct a one-layer convolutional neural net for the STL-10 dataset. Your network takes an input RGB image of dimensions $96 \times 96 \times 3$ and outputs a probability vector over the 10 classes. There are 16 convolutional filters followed by 16 pools in the pooling layer, and each convolutional filter is connected to a single pooling operation.

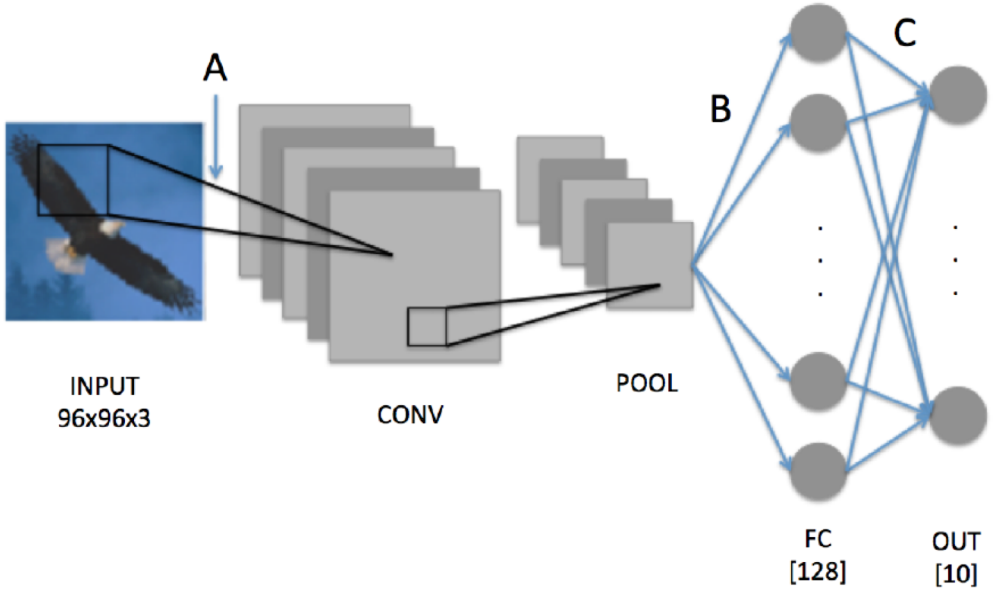


Figure 1: CNN for Problem 2

(a) (3 Points) Assume each of the 2D convolutional filters is of size 3×3. What is the shape of a single convolutional filter's weight matrix (label A)?

Spring 2017: Problem 2

This problem makes use of the STL-10 dataset, which contains 500 training images and 800 test images for each of 10 different classes: airplane, bird, car, cat, deer, dog, horse, monkey, ship, and truck. You decide to construct a one-layer convolutional neural net for the STL-10 dataset. Your network takes an input RGB image of dimensions $96 \times 96 \times 3$ and outputs a probability vector over the 10 classes. There are 16 convolutional filters followed by 16 pools in the pooling layer, and each convolutional filter is connected to a single pooling operation.

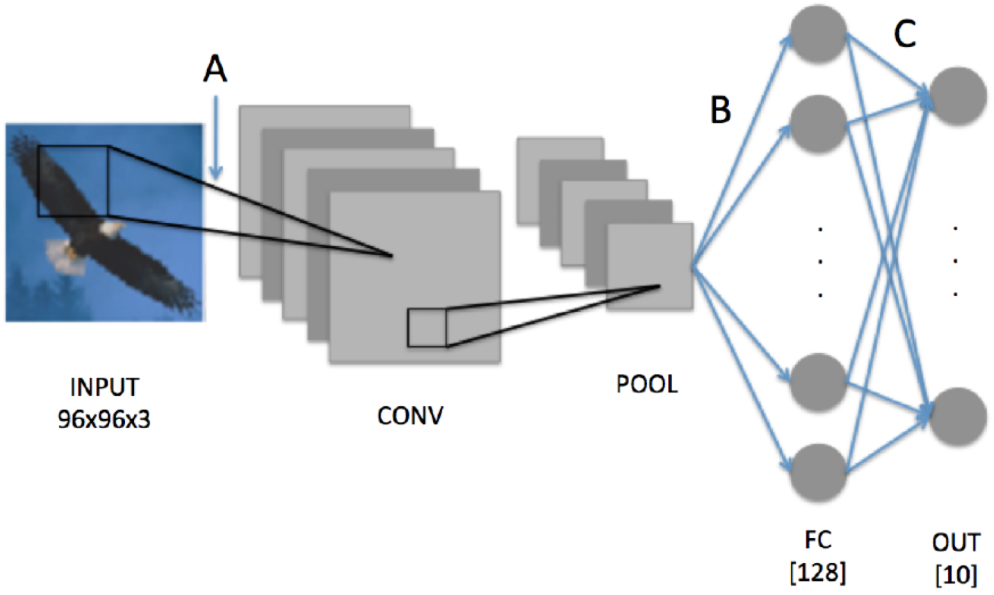


Figure 1: CNN for Problem 2

1 layer conv net

10 classes

Image: 96 x 96 x 3

16 conv filters

16 pools

(a) (3 Points) Assume each of the 2D convolutional filters is of size 3×3. What is the shape of a single convolutional filter's weight matrix (label A)?

Spring 2017: Problem 2

(a) (3 Points) Assume each of the 2D convolutional filters is of size 3×3. What is the shape of a single convolutional filter's weight matrix (label A)?

Spring 2017: Problem 2

(a) (3 Points) Assume each of the 2D convolutional filters is of size 3×3. What is the shape of a single convolutional filter's weight matrix (label A)?

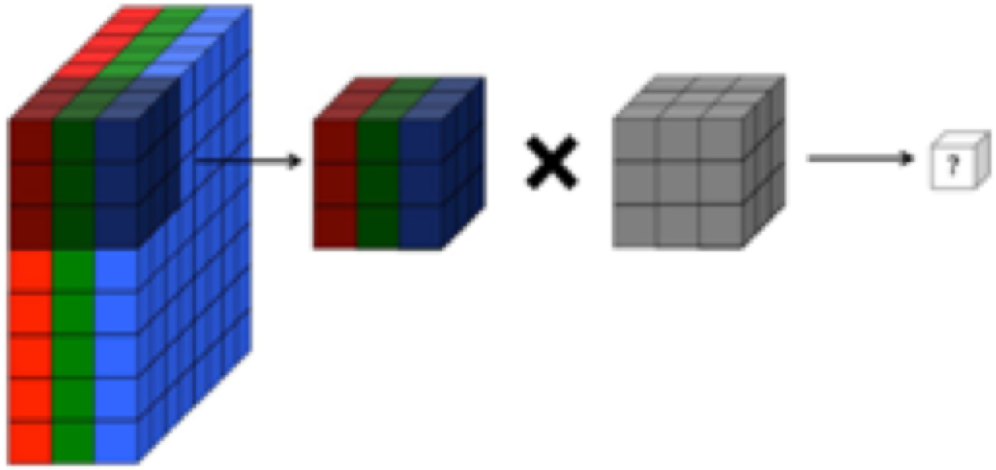


Figure 5-8. Representing a full-color RGB image as a volume and applying a volumetric convolutional filter

Image: 96 x 96 x 3

16 conv filters

https://ai.stackexchange.com/questions/5769/in-a-cnn-does-each-new-filter-have-different-weights-for-each-input-channel-or/5771

Piazza post @83, L03: conv neural nets

Spring 2017: Problem 2

(a) (3 Points) Assume each of the 2D convolutional filters is of size 3×3. What is the shape of a single convolutional filter's weight matrix (label A)?

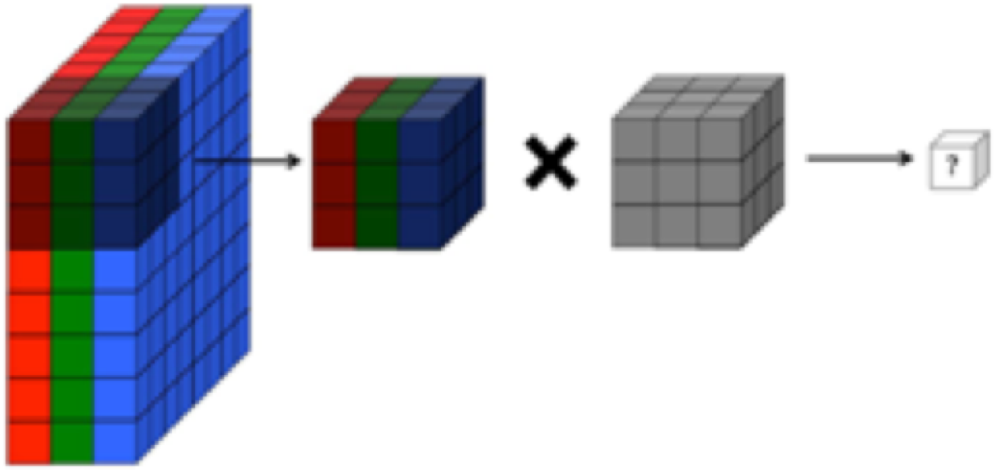


Figure 5-8. Representing a full-color RGB image as a volume and applying a volumetric convolutional filter

Image: 96 x 96 x 3

16 conv filters

Answer:

3 x 3 x 3 (need 3 channels for each RGB channel)

https://ai.stackexchange.com/questions/5769/in-a-cnn-does-each-new-filter-have-different-weights-for-each-input-channel-or/5771

Piazza post @83, L03: conv neural nets

Spring 2017: Problem 2

(b) (4 Points) What is the resulting shape of the output of a single convolutional filter? Assume all strides are 1, and there is no zero-padding

Spring 2017: Problem 2

(b) (4 Points) What is the resulting shape of the output of a single convolutional filter? Assume all strides are 1, and there is no zero-padding

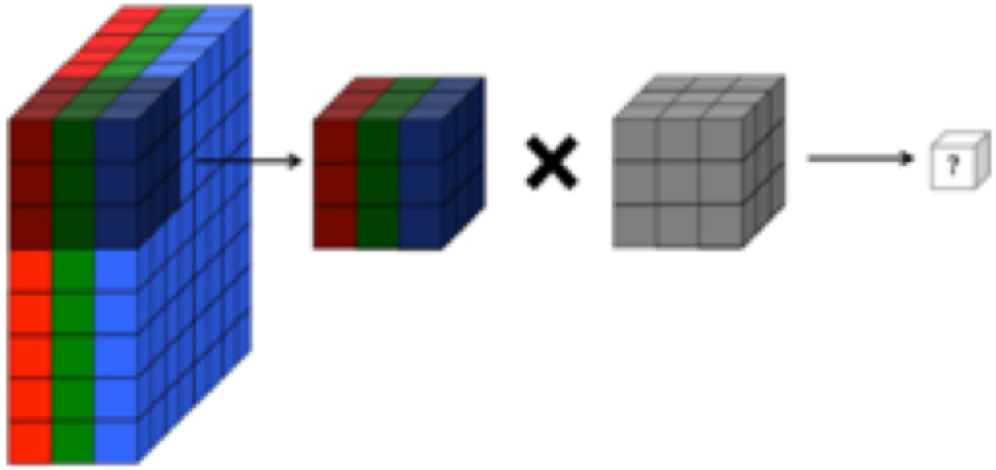


Figure 5-8. Representing a full-color RGB image as a volume and applying a volumetric convolutional filter

Image: 96 x 96 x 3
Filter: 3 x 3 x 3

NO zero-padding

https://ai.stackexchange.com/questions/5769/in-a-cnn-does-each-new-filter-have-different-weights-for-each-input-channel-or/5771

Piazza post @83, L03: conv neural nets

Spring 2017: Problem 2

(b) (4 Points) What is the resulting shape of the output of a single convolutional filter? Assume all strides are 1, and there is no zero-padding

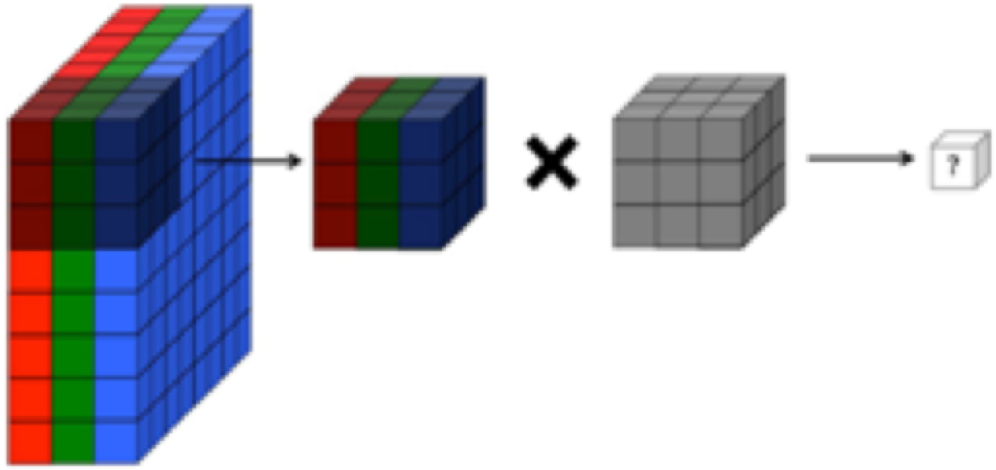


Figure 5-8. Representing a full-color RGB image as a volume and applying a volumetric convolutional filter

Image: 96 x 96 x 3
Filter: 3 x 3 x 3

NO zero-padding

Answer:
94 x 94

https://ai.stackexchange.com/questions/5769/in-a-cnn-does-each-new-filter-have-different-weights-for-each-input-channel-or/5771

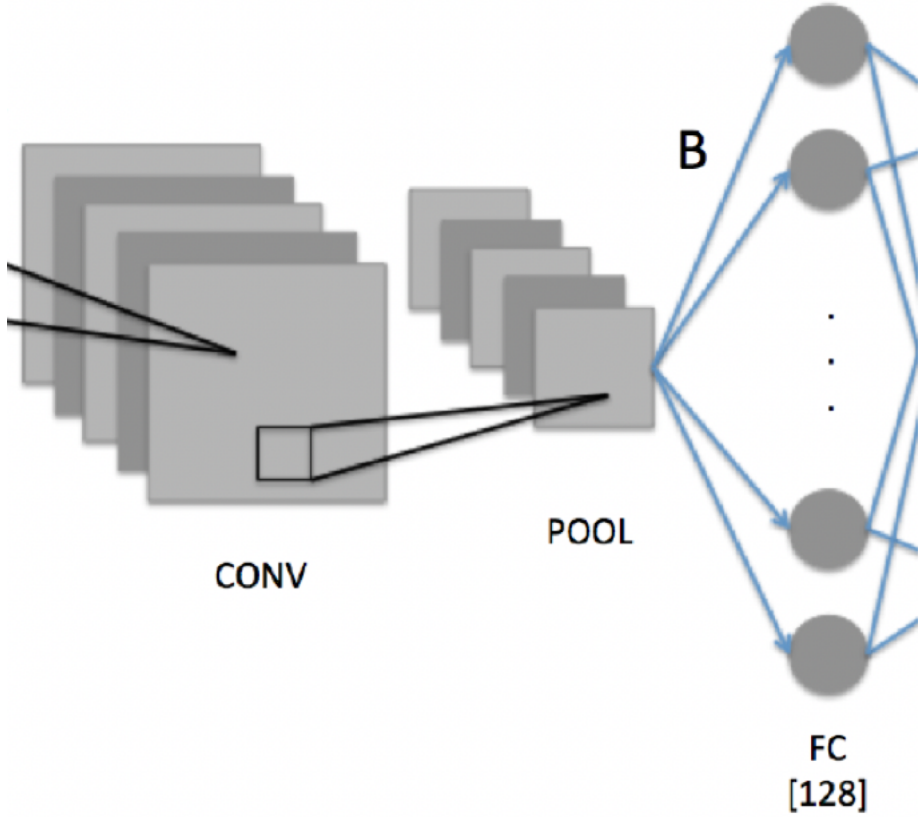
Piazza post @83, L03: conv neural nets

Spring 2017: Problem 2

(c) (4 Points) You are employing a 2×2 max pooling layer with a stride of 2 in both directions with no zero-padding. What is the shape of the weight matrix that connects the output of all 16 pooling units in the pool layer to the 128 units in the fully connected layer (labeled B)?

Spring 2017: Problem 2

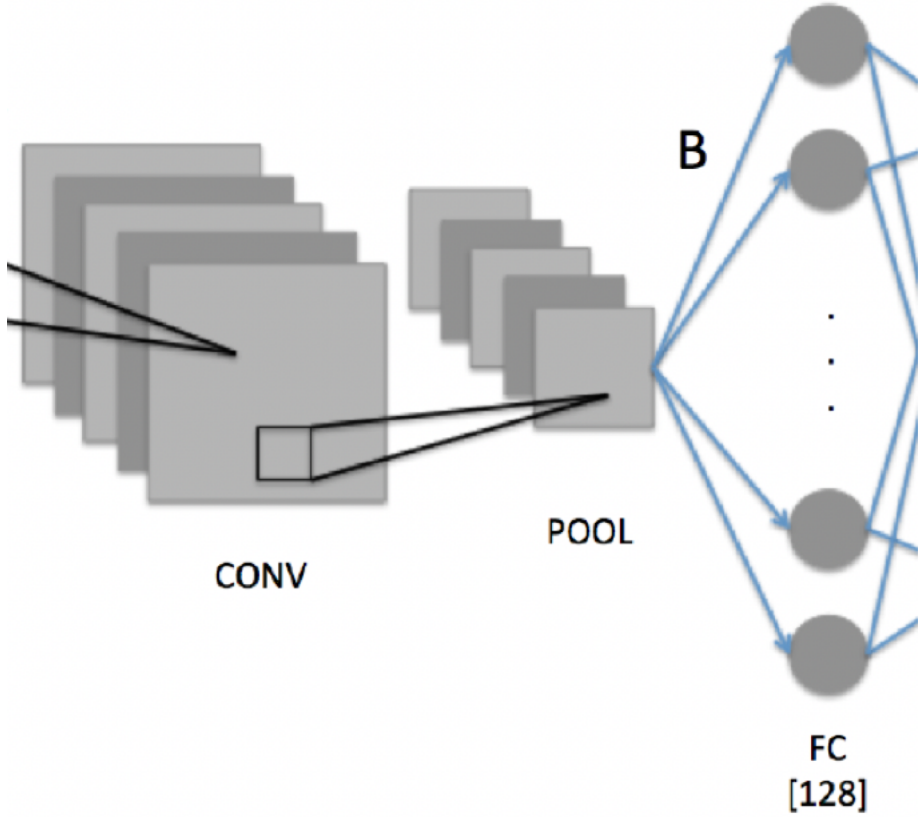
(c) (4 Points) You are employing a 2×2 max pooling layer with a stride of 2 in both directions with no zero-padding. What is the shape of the weight matrix that connects the output of all 16 pooling units in the pool layer to the 128 units in the fully connected layer (labeled B)?



Conv Output: 94 x 94 x 16
Pool Output: 47 x 47 x 16
Next Layer: 128

Spring 2017: Problem 2

(c) (4 Points) You are employing a 2×2 max pooling layer with a stride of 2 in both directions with no zero-padding. What is the shape of the weight matrix that connects the output of all 16 pooling units in the pool layer to the 128 units in the fully connected layer (labeled B)?



Conv Output: 94 x 94 x 16
Pool Output: 47 x 47 x 16
Next Layer: 128

Answer:
47 x 47 x 16 x 128

Spring 2017: Problem 2

(d) (3 Points) What is the shape of the weight matrix between the two fully connected layers (labeled C)?

FC Layer 1: 128
FC Layer 2: 10

Spring 2017: Problem 2

(d) (3 Points) What is the shape of the weight matrix between the two fully connected layers (labeled C)?

FC Layer 1: 128
FC Layer 2: 10

Answer:
128 x 10

Spring 2017: Problem 2

(d) (3 Points) What is the shape of the weight matrix between the two fully connected layers (labeled C)?

FC Layer 1: 128
FC Layer 2: 10

Answer:
128 x 10

(e) (3 Points) What activation function of the final 10 outputs will you need to use to produce a vector of probabilities that sum to 1?

Spring 2017: Problem 2

(d) (3 Points) What is the shape of the weight matrix between the two fully connected layers (labeled C)?

FC Layer 1: 128
FC Layer 2: 10

Answer:
128 x 10

(e) (3 Points) What activation function of the final 10 outputs will you need to use to produce a vector of probabilities that sum to 1?

Answer:
softmax

Spring 2017: Problem 3

Consider a simplified recurrent neural network architecture that operates on 1-D inputs $x_t \in \mathbb{R}$ and updates a 1-D hidden state $h_t \in \mathbb{R}$ at each time-step $t = 1, \ldots, T$ as follows:

$$h_t = v \cdot x_t + w \cdot h_{t-1} \tag{1}$$

where the RNN-parameters $w, v \in \mathbb{R}$ are both simply scalars rather than vectors/matrices. Each training sequence of inputs, denoted as $(x_1, \ldots, x_T)$, will be associated with a single scalar label $y \in \mathbb{R}$. To predict y given $(x_1, \ldots, x_T)$, we will simply use the final RNN hidden state h_T (at time-step T corresponding to the final input), and our prediction will incur mean-squared error loss $(y - h_T)^2$. For this problem, you *must* always assume each input $x_t \in \{0, 1\}$ takes either value 0 or 1, and we always set the initial hidden state $h_0 = 0$ before the RNN operates on a given input sequence.

$h_t = v*x_t + w*h_{t-1}$

Loss = $(y - h_T)^2$

$h_0 = 0$

Spring 2017: Problem 3

a) (4 Points) Is there a setting of the parameters w, v such that this RNN will output a predicted value $h_T = 1$ when given the input sequence (of length $T = 7$): (0,0,0,0,0,0,0) where each $x_t = 0$ for $t = 1, \ldots, 7$? If yes, specify the values of w, v that would result in this prediction. Otherwise, explain why no such parameter-values exist.

$h_T = 1$

$h_0 = 0$

$h_t = v * x_t + w * h_{t-1}$

x = (0,0,0…,0)

Spring 2017: Problem 3

a) (4 Points) Is there a setting of the parameters w, v such that this RNN will output a predicted value $h_T = 1$ when given the input sequence (of length $T = 7$): (0,0,0,0,0,0,0) where each $x_t = 0$ for $t = 1, \ldots, 7$? If yes, specify the values of w, v that would result in this prediction. Otherwise, explain why no such parameter-values exist.

$h_T = 1$

$h_0 = 0$

$h_t = v * x_t + w * h_{t-1}$

$x = (0,0,0\ldots,0)$

$h_0 = 0$

$h_1 = v * x_1 + w * h_0$

$h_1 = v * 0 + w * 0 = 0$

$h_2 = v * x_2 + w * h_1$

$h_2 = v * 0 + w * 0 = 0$

$h_t = 0$ for all t!

Spring 2017: Problem 3

a) (4 Points) Is there a setting of the parameters w, v such that this RNN will output a predicted value $h_T = 1$ when given the input sequence (of length $T = 7$): (0,0,0,0,0,0,0) where each $x_t = 0$ for $t = 1, \ldots, 7$? If yes, specify the values of w, v that would result in this prediction. Otherwise, explain why no such parameter-values exist.

$h_T = 1$

$h_0 = 0$

$h_t = v * x_t + w * h_{t-1}$

$x = (0,0,0\ldots,0)$

Answer:

no parameters exist

$h_0 = 0$

$h_1 = v * x_1 + w * h_0$

$h_1 = v * 0 + w * 0 = 0$

$h_2 = v * x_2 + w * h_1$

$h_2 = v * 0 + w * 0 = 0$

$h_t = 0$ for all t!

Spring 2017: Problem 3

b) (4 Points) Suppose we have a training set of (sequence, label) pairs where we always have $y = 2 \cdot x_T$ (the label associated with the sequence is always twice the value of the input at the final position). Is there a setting of the parameters w, v such that this RNN model could achieve zero training loss on all datasets (of arbitrary sample-size)? If yes, specify the values of w, v. Otherwise, provide an example of dataset (you are free to choose the sample-size and sequence-length T) where our RNN would not be able to achieve zero training loss.

$y = 2x_T$

$h_0 = 0$

$h_t = v * x_t + w * h_{t-1}$

Spring 2017: Problem 3

b) (4 Points) Suppose we have a training set of (sequence, label) pairs where we always have $y = 2 \cdot x_T$ (the label associated with the sequence is always twice the value of the input at the final position). Is there a setting of the parameters w, v such that this RNN model could achieve zero training loss on all datasets (of arbitrary sample-size)? If yes, specify the values of w, v. Otherwise, provide an example of dataset (you are free to choose the sample-size and sequence-length T) where our RNN would not be able to achieve zero training loss.

$y = 2x_T$

$h_0 = 0$

$h_t = v * x_t + w * h_{t-1}$

$h_1 = v * x_1 + w * h_0$

$h_1 = v * x_1 + w * 0$

$h_1 = v * x_1$

$h_1 = 2x_T$

$h_2 = v * x_2 + w * h_1$

$h_2 = v * x_2 + 0 * h_1$

$h_2 = v * x_2$

$h_2 = 2x_T$

Spring 2017: Problem 3

b) (4 Points) Suppose we have a training set of (sequence, label) pairs where we always have $y = 2 \cdot x_T$ (the label associated with the sequence is always twice the value of the input at the final position). Is there a setting of the parameters w, v such that this RNN model could achieve zero training loss on all datasets (of arbitrary sample-size)? If yes, specify the values of w, v. Otherwise, provide an example of dataset (you are free to choose the sample-size and sequence-length T) where our RNN would not be able to achieve zero training loss.

$y = 2x_T$

$h_0 = 0$

$h_t = v * x_t + w * h_{t-1}$

Answer:

v = 2 ; w = 0

$h_1 = v * x_1 + w * h_0$

$h_1 = v * x_1 + w * 0$

$h_1 = v * x_1$

$h_1 = 2x_T$

$h_2 = v * x_2 + w * h_1$

$h_2 = v * x_2 + 0 * h_1$

$h_2 = v * x_2$

$h_2 = 2x_T$

Spring 2017: Problem 4

a) (2 Points) Given a first-layer convolutional filter from a CNN trained to predict TF binding from DNA sequences, we can get the nucleotide probability at each position by normalizing each column of the filter to sum to one ? (Yes / No)

b) (2 Points) We can select the best set of hyper-parameters by finding the combination that gives the lowest training loss? (Yes / No)

Spring 2017: Problem 4

a) (2 Points) Given a first-layer convolutional filter from a CNN trained to predict TF binding from DNA sequences, we can get the nucleotide probability at each position by normalizing each column of the filter to sum to one ? (Yes / No)

b) (2 Points) We can select the best set of hyper-parameters by finding the combination that gives the lowest training loss? (Yes / No)

Answer:

a) No – weight values not tied to nt frequency

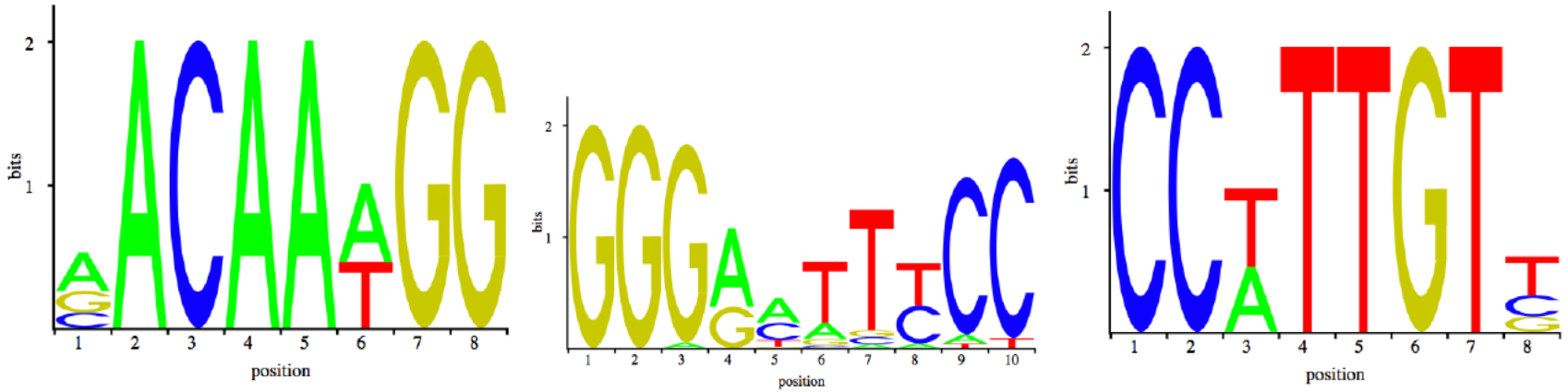
b) No – lowest validation loss

Spring 2017: Problem 4

Suppose we have the following convolutional kernel from a CNN trained to predict TF binding from DNA sequences.

$$\begin{bmatrix} A \\ C \\ G \\ T \end{bmatrix} = \begin{bmatrix} -1.1 & -0.5 & 1.3 & 0.3 & 0.9 & 0.19 & 0.09 & 0.22 & -0.09 & 0.01 \\ 4.1 & 2.5 & -0.3 & -0.08 & 1.3 & -0.13 & -0.05 & 0.36 & 0.03 & 0.1 \\ 0.5 & 0.1 & -0.1 & 0.13 & 0.01 & 0.91 & 0.13 & 0.11 & 0.01 & -0.01 \\ 0.6 & 0.5 & 1.6 & 2.5 & 3.1 & -0.57 & 0.73 & 0.21 & -0.05 & 0.12 \end{bmatrix}$$

And we have the forward-strand motifs of the following TFs represented in bit-information logo as follows:



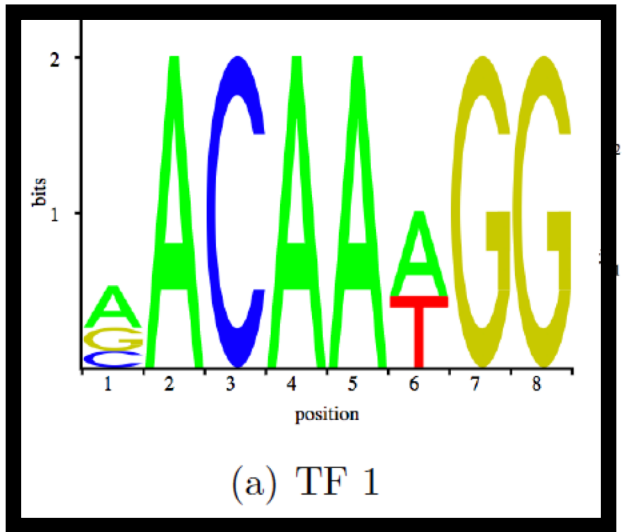
(a) TF 1 (b) TF 2 (c) TF 3

Spring 2017: Problem 4

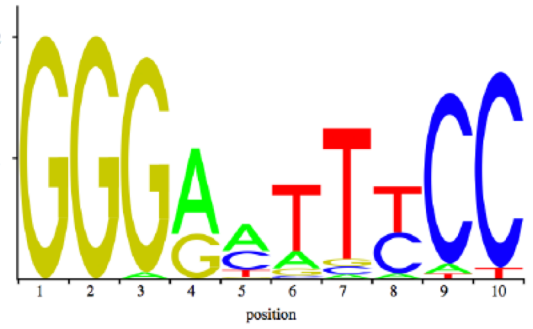
Suppose we have the following convolutional kernel from a CNN trained to predict TF binding from DNA sequences.

$$\begin{bmatrix} A \\ C \\ G \\ T \end{bmatrix} = \begin{bmatrix} -1.1 & -0.5 & 1.3 & 0.3 & 0.9 & 0.19 & 0.09 & 0.22 & -0.09 & 0.01 \\ 4.1 & 2.5 & -0.3 & -0.08 & 1.3 & -0.13 & -0.05 & 0.36 & 0.03 & 0.1 \\ 0.5 & 0.1 & -0.1 & 0.13 & 0.01 & 0.91 & 0.13 & 0.11 & 0.01 & -0.01 \\ 0.6 & 0.5 & 1.6 & 2.5 & 3.1 & -0.57 & 0.73 & 0.21 & -0.05 & 0.12 \end{bmatrix}$$

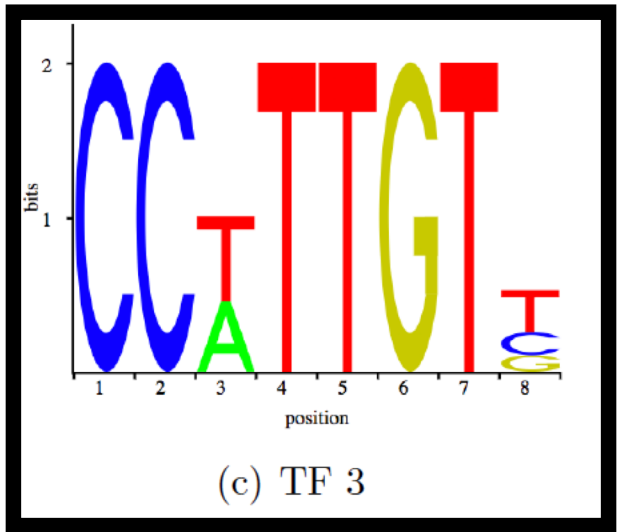
And we have the forward-strand motifs of the following TFs represented in bit-information logo as follows:



(a) TF 1



(b) TF 2



(c) TF 3

Spring 2017: Problem 4

One way to interpret a specific neuron in a neural network is to fix the network weights, and find the input that can maximize the output of that neuron. Consider a one-layer linear network with weight $W = [w_1, w_2, w_3]^T$. Suppose the input is $X = [x_1, x_2, x_3]^T$, then the output is $y = W^T X = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$. In the following questions, assume W is fixed and $W = [1, 2, 3]^T$.

d) (2 Points) Is there a finite input X that can maximize Y? If yes, specify the optimal input vector. If not, explain why.

Spring 2017: Problem 4

One way to interpret a specific neuron in a neural network is to fix the network weights, and find the input that can maximize the output of that neuron. Consider a one-layer linear network with weight $W = [w_1, w_2, w_3]^T$. Suppose the input is $X = [x_1, x_2, x_3]^T$, then the output is $y = W^T X = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$. In the following questions, assume W is fixed and $W = [1, 2, 3]^T$.

d) (2 Points) Is there a finite input X that can maximize Y? If yes, specify the optimal input vector. If not, explain why.

$y = W^T X$

Spring 2017: Problem 4

One way to interpret a specific neuron in a neural network is to fix the network weights, and find the input that can maximize the output of that neuron. Consider a one-layer linear network with weight $W = [w_1, w_2, w_3]^T$. Suppose the input is $X = [x_1, x_2, x_3]^T$, then the output is $y = W^T X = w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$. In the following questions, assume W is fixed and $W = [1, 2, 3]^T$.

d) (2 Points) Is there a finite input X that can maximize Y? If yes, specify the optimal input vector. If not, explain why.

$y = W^T X$

Answer:

No – if X → infinity, Y → infinity

Spring 2017: Problem 4

e) (4 Points) If we force the L_2 norm of X to be 1, i.e. $(x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}} = 1$, is there a finite input X that can maximize Y? If yes, specify the optimal input vector. If not, explain why.

Spring 2017: Problem 4

e) (4 Points) If we force the L_2 norm of X to be 1, i.e. $(x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}} = 1$, is there a finite input X that can maximize Y? If yes, specify the optimal input vector. If not, explain why.

$Y = w_1 x_1 + w_2 x_2 + w_3 x_3$

$Y = x_1 + 2x_2 + 3x_3$ $\quad 1 = x_1^2 + x_2^2 + x_3^2$

$Y = x_1 + 2x_2 + 3\sqrt{1 - x_1^2 - x_2^2}$

Take $\frac{dY}{dx_1}$ set $= 0$ find x_1^*

$\frac{dY}{dx_1} = 0 = 1 + 2\frac{dx_2}{dx_1}^{\to 0} + 3 \cdot \frac{1}{2}(1 - x_1^2 - x_2^2)^{-1/2} \cdot (-2x_1)$ be orthogonal

$0 = 1 - 3x_1(1 - x_1^2 - x_2^2)^{-1/2}$

$3x_1 = \sqrt{1 - x_1^2 - x_2^2}$

Take $\frac{dY}{dx_2}$ set $= 0$ find x_2^*

$\frac{dY}{dx_2} = 0 = \frac{dx_1}{dx_2}^{\to 0} + 2 + 3 \cdot \frac{1}{2}(1 - x_1^2 - x_2^2)^{-1/2}(-2x_2)$

$0 = 2 - 3x_2(1 - x_1^2 - x_2^2)^{-1/2}$

$\frac{3}{2}x_2 = \sqrt{1 - x_1^2 - x_2^2}$

So then we algebra to get

$3x_1 = \frac{3}{2}x_2 \Rightarrow x_2 = 2x_1$

$3x_1 = \sqrt{1 - x_1^2 - (2x_1)^2} \Rightarrow x_1^* = \frac{1}{\sqrt{14}}$ $\quad x_2^* = \frac{2}{\sqrt{14}}$

$x_3 = \sqrt{1 - x_1^2 - x_2^2}$

$x_3^* = \frac{3}{\sqrt{14}}$

Onto recitation R10!

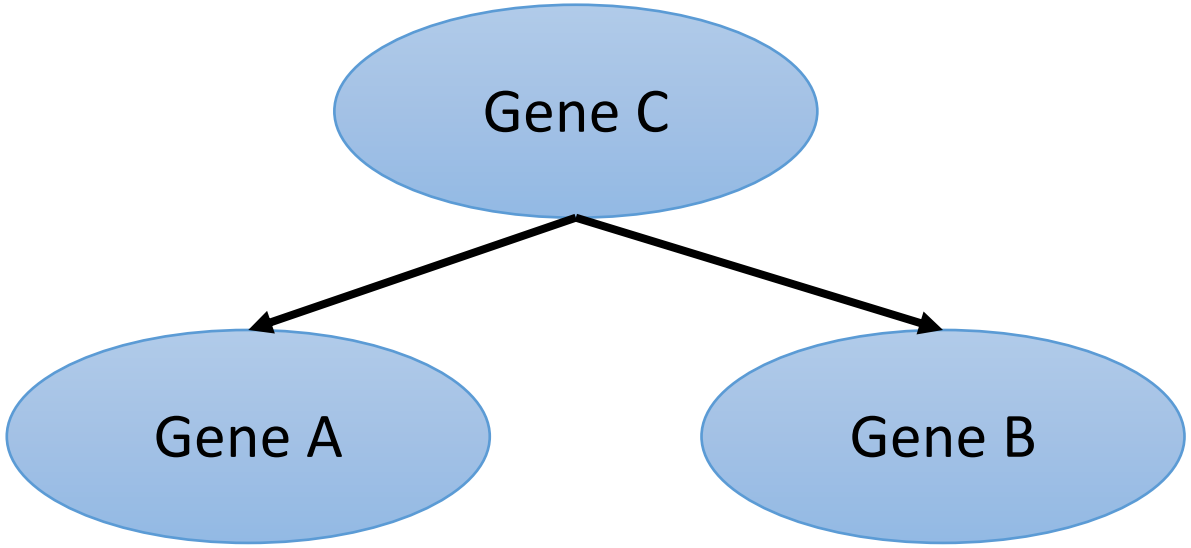
A. Spring 2017
   - Problem 2a-e
   - Problem 3a,b
   - Problem 4a-e

B. Spring 2019
   - Problem 1a-c

C. Spring 2020
   - Problem 4a-d

Spring 2019: Problem 1

a) (5 Points) You observe that Gene A and Gene B are highly correlated in expression data for 20 different conditions. Whenever Gene A is expressed Gene B is expressed. To investigate further you knock out Gene A, and note that Gene B is still expressed in certain conditions. Explain how this could be happening given your data was consistent with the causal regulation of Gene B by Gene A.



Spring 2019: Problem 1

b) (6 Points) You hypothesize that five transcription factors $TF_1, \ldots TF_5$ may regulate the transcription of gene G. You perform intervention experiments on the transcription factors one at a time, and calculate on an individual basis the probability that you would observe the changes in gene G transcripts in your data at random (null hypothesis) for $TF_1, \ldots TF_5$ as 0.003, 0.006, 0.020, 0.045, and 0.600, respectively. Using Bonferonni multiple hypothesis correction for which TFs can you reject the null hypothesis? Using Benjamini-Hochberg for an expected false discovery rate (FDR) of 0.05, for which TFs can you reject the null hypothesis that the observed expression change is occurring at random?

TF	Raw P-val
1	0.003
2	0.006
3	0.020
4	0.045
5	0.600

Spring 2019: Problem 1

b) (6 Points) You hypothesize that five transcription factors $TF_1, \ldots TF_5$ may regulate the transcription of gene G. You perform intervention experiments on the transcription factors one at a time, and calculate on an individual basis the probability that you would observe the changes in gene G transcripts in your data at random (null hypothesis) for $TF_1, \ldots TF_5$ as 0.003, 0.006, 0.020, 0.045, and 0.600, respectively. Using Bonferonni multiple hypothesis correction for which TFs can you reject the null hypothesis? Using Benjamini-Hochberg for an expected false discovery rate (FDR) of 0.05, for which TFs can you reject the null hypothesis that the observed expression change is occurring at random?

TF	Raw P-val	< 0.01?
1	0.003	Y
2	0.006	Y
3	0.020	N
4	0.045	N
5	0.600	N

Bonferroni threshold :

< 0.05 / 5
< 0.01

Spring 2019: Problem 1

b) (6 Points) You hypothesize that five transcription factors $TF_1, \ldots TF_5$ may regulate the transcription of gene G. You perform intervention experiments on the transcription factors one at a time, and calculate on an individual basis the probability that you would observe the changes in gene G transcripts in your data at random (null hypothesis) for $TF_1, \ldots TF_5$ as 0.003, 0.006, 0.020, 0.045, and 0.600, respectively. Using Bonferonni multiple hypothesis correction for which TFs can you reject the null hypothesis? Using Benjamini-Hochberg for an expected false discovery rate (FDR) of 0.05, for which TFs can you reject the null hypothesis that the observed expression change is occurring at random?

TF	Raw P-val	Rank	B-H thresh	< thresh?
1	0.003	1	0.01	Y
2	0.006	2	0.02	Y
3	0.020	3	0.03	Y
4	0.045	4	0.04	N
5	0.600	5	0.05	N

B-H threshold :

< 0.05 * (i / m)
< 0.05 * (i / 5)
< 0.01 * i

Spring 2019: Problem 1

c) (6 Points) We have observed we can regularize weight values by adding a term onto a loss function that penalizes weight magnitude. In Figure 1 below we show $\|w\|\|_1$ (L1 norm, left) and $\|w\|_2$ (L2 norm, right). The point w* in both figures is on the same loss contour but makes different weight choices that minimize the respective norm. Describe which method will choose the sparsest set of non-zero weights and why.

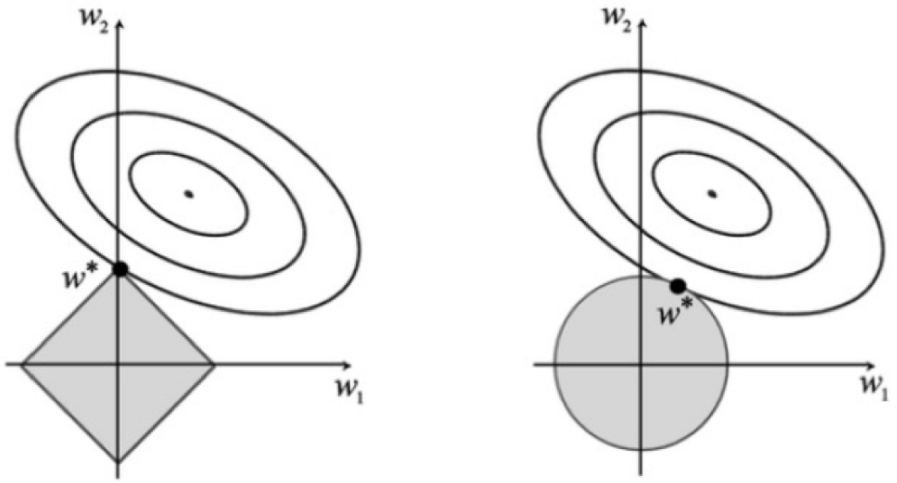


Figure 1: Weight regularization at a fixed loss contour

Spring 2019: Problem 1

c) (6 Points) We have observed we can regularize weight values by adding a term onto a loss function that penalizes weight magnitude. In Figure 1 below we show $\|w\|_1$ (L1 norm, left) and $\|w\|_2$ (L2 norm, right). The point w* in both figures is on the same loss contour but makes different weight choices that minimize the respective norm. Describe which method will choose the sparsest set of non-zero weights and why.

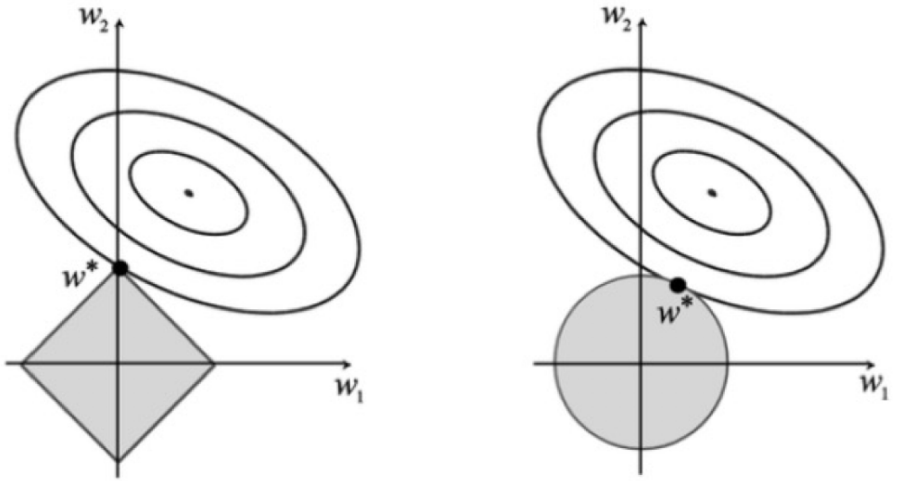
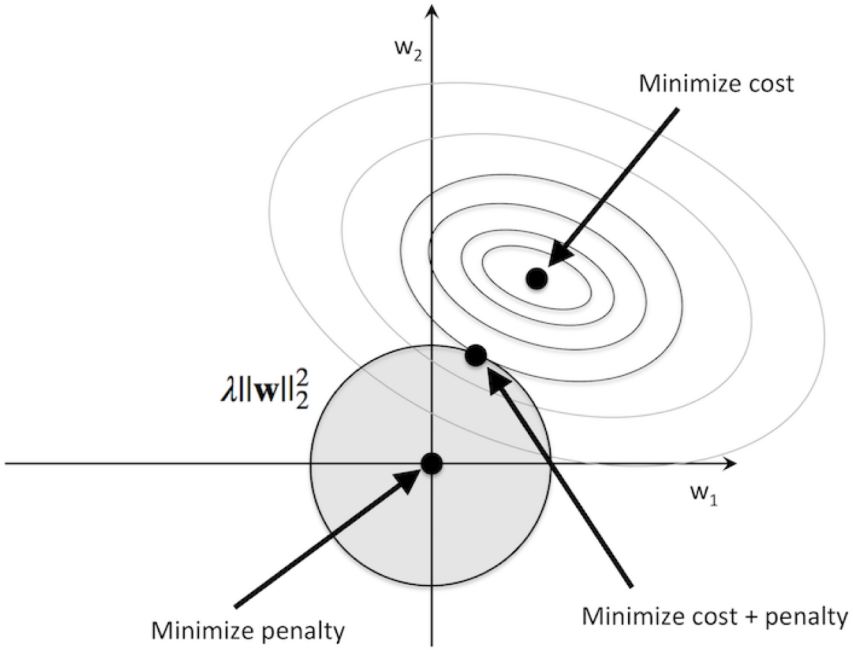


Figure 1: Weight regularization at a fixed loss contour



$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \qquad \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

Spring 2019: Problem 1

c) (6 Points) We have observed we can regularize weight values by adding a term onto a loss function that penalizes weight magnitude. In Figure 1 below we show $\|w\|_1$ (L1 norm, left) and $\|w\|_2$ (L2 norm, right). The point w* in both figures is on the same loss contour but makes different weight choices that minimize the respective norm. Describe which method will choose the sparsest set of non-zero weights and why.

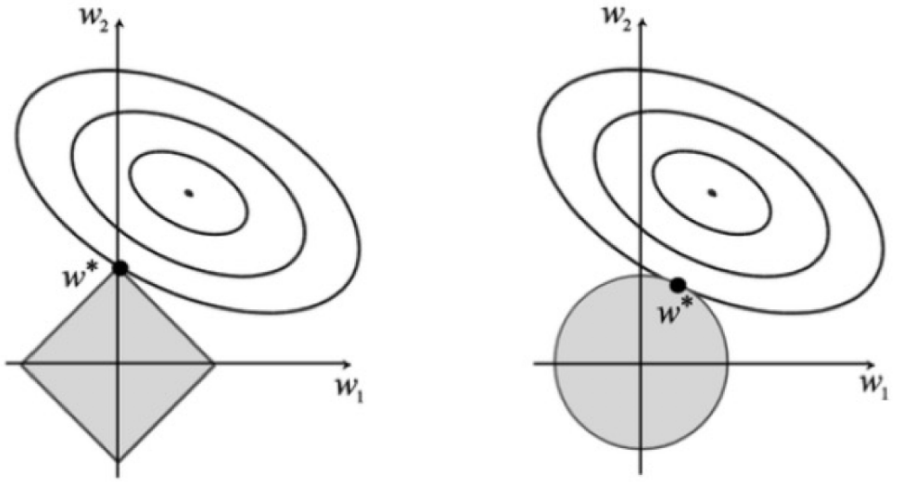
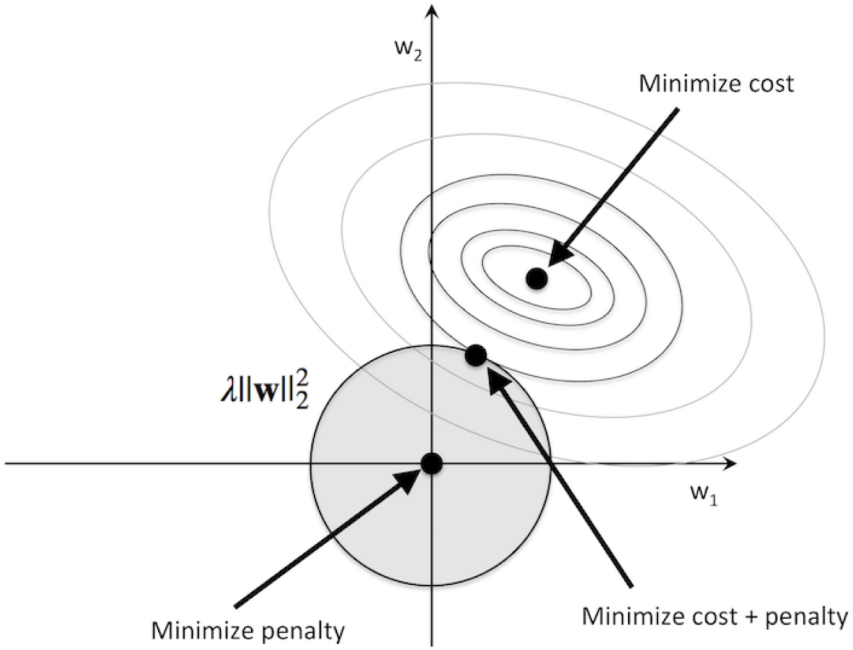


Figure 1: Weight regularization at a fixed loss contour



$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \qquad \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

Answer:

L1 norm → sends weights to zero

Onto recitation R10!

A. Spring 2017
   Problem 2a-e
   Problem 3a,b
   Problem 4a-e

B. Spring 2019
   Problem 1a-c

C. Spring 2020
   Problem 4a-d

Spring 2020: Problem 4

Your colleague has just completed a whole-genome sequencing study of 50,000 Schizophrenia cases and 100,000 controls, and is enlisting your help in interpreting the data. *Note: In your responses to the following questions, feel free to draw figures, use bulleted lists, etc. Primarily, we are looking for a well-reasoned response to these open-ended questions.*

a) (5 points) You first focus on common variants associated with Schizophrenia and seek to predict their functional impact. You decide to focus on their impact on enhancer activity and transcription factor binding. Why does it make sense to focus on enhancer activity and TF binding for common variants?

Spring 2020: Problem 4

Your colleague has just completed a whole-genome sequencing study of 50,000 Schizophrenia cases and 100,000 controls, and is enlisting your help in interpreting the data. *Note: In your responses to the following questions, feel free to draw figures, use bulleted lists, etc. Primarily, we are looking for a well-reasoned response to these open-ended questions.*

a) (5 points) You first focus on common variants associated with Schizophrenia and seek to predict their functional impact. You decide to focus on their impact on enhancer activity and transcription factor binding. Why does it make sense to focus on enhancer activity and TF binding for common variants?

- Common variants == weak effects; rare variants == strong effects

- Weak effects are selected for in enhancers (enhancer activity), non-coding regions (TF binding sites), regulatory variants, DNAse regions (accessible chromatin)

Spring 2020: Problem 4

b) (7 points) (i) What deep learning architectures and features can you use to prioritize variants that are more likely to be functional? Does the training data need to be Schizophrenia-specific? (ii) What input features would you use, and what datasets can you use to train your model? (iii) Would your model also lead to meaningful results for rare non-coding variants?

Spring 2020: Problem 4

b) (7 points) (i) What deep learning architectures and features can you use to prioritize variants that are more likely to be functional? Does the training data need to be Schizophrenia-specific? (ii) What input features would you use, and what datasets can you use to train your model? (iii) Would your model also lead to meaningful results for rare non-coding variants?

-lots of architecture options, but likely CNN best choice

-input would be one hot encoded DNA (think PSet 2)
    -features of input:
    motifs (discovered from model training on your data)
    hardcoded motifs (from literature/prior knowledge)
    grammar (motif spacing)

-datasets: ChIP-seq, Dnase accessible site, bound vs non bound for TFs
-no need for dataset to be schizophrenia specific but helps if brain-specific

-meaningful for non-coding variants since operating on the DNA seq

Spring 2020: Problem 4

c) (7 points) You now turn your attention to finding the impact of genetic variation on protein-coding gene function. (i) How can you leverage your work on non-coding variants to help with this task? Hint: will eQTLs be helpful? What about Hi-C information? What about splicing QTLs? Please be specific about the data sources and tools you would combine to predict impact on protein-coding gene function. (ii) How would you combine common and rare variants together to prioritize candidate target genes and their function? Describe a deep learning architecture or other machine learning model in detail for achieving this.

Spring 2020: Problem 4

c) (7 points) You now turn your attention to finding the impact of genetic variation on protein-coding gene function. (i) How can you leverage your work on non-coding variants to help with this task? Hint: will eQTLs be helpful? What about Hi-C information? What about splicing QTLs? Please be specific about the data sources and tools you would combine to predict impact on protein-coding gene function. (ii) How would you combine common and rare variants together to prioritize candidate target genes and their function? Describe a deep learning architecture or other machine learning model in detail for achieving this.

- Map non-coding variants to their target genes (using Hi-C, etc.)

- Common variants == weak effects; rare variants == strong effects
- Common variants == non-coding; rare variants == coding

- Need to combine common + rare variants (lots of ways to do this!)

Spring 2020: Problem 4

d) (6 points) (i) How would you visualize the genes you predicted in a tSNE plot of single-cell gene expression data in brain to recognize the cell types where your prioritized genes act? (ii) Describe an empirical-prior Bayesian framework for prioritizing candidate driver genes using the cell-type-specific enrichments you found.

Spring 2020: Problem 4

d) (6 points) (i) How would you visualize the genes you predicted in a tSNE plot of single-cell gene expression data in brain to recognize the cell types where your prioritized genes act? (ii) Describe an empirical-prior Bayesian framework for prioritizing candidate driver genes using the cell-type-specific enrichments you found.

- tSNE plot of single cell gene expression data --> clusters of cell types
- Color / superimpose the expression level of some genes on top of tSNE
- Prioritize gene candidates – how do you find some genes that are enriched in certain clusters of cells? (without having to try every gene)