Consider 3 weight vectors with 3 dimensions:

w_1 = <1.0, -2.0, 0.5>

w_2 = <0.1, -0.2, 0.05>

w_3 = <100.0, -200.0, 50.0>

On a training dataset, they give the following errors:

$L(\mathbf{w_1}) = 10.0$

$L(\mathbf{w_2}) = 100.0$

$L(\mathbf{w_3}) = 8.0$

1. Calculate the L2 norm of each weight vector.

$||\mathbf{w_1}|| = ?$

$||\mathbf{w_2}|| = ?$

$||\mathbf{w_3}|| = ?$

1. Calculate the L2 norm of each weight vector.

$||\mathbf{w}_1||$ = sqrt((1)^2 + (-2)^2 + (.5)^2) = 2.29

$||\mathbf{w}_2||$ = sqrt((.1)^2 + (-.2)^2 + (.05)^2) = .229

$||\mathbf{w}_3||$ = sqrt((100)^2 + (-200)^2 + (50)^2) = 229

Assume you apply L2 regularization, where you minimize the combined function:

$L(\mathbf{w}) + \lambda R(\mathbf{w})$, where $R(\mathbf{w}) = ||\mathbf{w}||$.

(Note: more often, the *squared* L2 norm is used for L2 regularization, but here let's use the L2 norm without squaring it).

2. For each of the 3 weight vectors, and each of the following values of λ, calculate the value of the function, $L(\mathbf{w}) + \lambda R(\mathbf{w})$.

	λ=0.001	λ=0.1	λ=100.0
$\mathbf{w}_1$	?	?	?
$\mathbf{w}_2$	?	?	?
$\mathbf{w}_3$	?	?	?

2. For each of the 3 weight vectors, and each of the following values of λ, calculate the value of the function, $L(\mathbf{w}) + \lambda R(\mathbf{w})$.

	λ=0.001	λ=0.1	λ=100.0
$\mathbf{w}_1$	10+.003=10.0023	10+.229=10.229	10+229=239
$\mathbf{w}_2$	100+.0002=100.00	100+.023=100.023	100+22.9=122.9
$\mathbf{w}_3$	8+.229=8.229	8+22.9 = 30.9	8+22900=22908

3. Based on your answer to Question 2, for each value of λ, say which weight vector is optimal.

For each one, say whether you think it is overfitting, underfitting, or neither. (There is no single best answer here.)

λ=.001: ?

λ=0.1: ?

λ=100: ?

3. Based on your answer to Question 2, for each value of λ, say which weight vector is optimal.

For each one, say whether you think it is overfitting, underfitting, or neither. (There is no single best answer here.)

λ=.001: $\mathbf{w}_3$ is the best vector. However, it may be overfitting, because the weights are much larger than $\mathbf{w}_1$, while only providing a small benefit in training loss (8 vs 10).

λ=0.1: $\mathbf{w}_1$ is the best vector, which seems to provide a good tradeoff between having small weights and also small loss.

λ=100: $\mathbf{w}_2$ is the best vector, even though it has poor loss, because it has the smallest norm. The poor loss suggests this model is underfitting.

3. Assume you are using an SVM. Calculate the margin for each weight vector.

Margin with $\mathbf{w_1}$: ?

Margin with $\mathbf{w_2}$: ?

Margin with $\mathbf{w_3}$: ?

4. Assume you are using an SVM. Calculate the margin for each weight vector.

Margin with $\mathbf{w_1}$: 2 / 2.29 = .873

Margin with $\mathbf{w_2}$: 2 / .229 = 8.73

Margin with $\mathbf{w_3}$: 2 / 229 = .00873

Using the formula for the margin, $2 / ||\mathbf{w}||$.

5. Discuss how adjusting λ adjusts the tradeoff of bias and variance.

5. Discuss how adjusting λ adjusts the tradeoff of bias and variance.

Increasing λ increases the bias toward having small weights. This reduces variance (the weights will be similarly small on different training sets) at the expense of having a high loss.

Decreasing λ increases the variance, in that the weights are allowed to fit the training data as closely as possible, even if it means the weights need to be large to do that. There is less bias toward small weights because R(**w**) is only a small part of the function when λ is small.

In general, high variance is associated with overfitting and high bias is associated with underfitting.