Congratulations! You passed!

TO PASS 80% or higher

Keep Learning

GRADE
100%

The Regression side of Supervised Learning

LATEST SUBMISSION GRADE
100%

1. What is the hypothesis space of linear regression? **1 / 1 point**

- [x] The set of flat hyperplanes

> **Correct**
> Correct, in general linear regression considers all possible flat planes in the appropriate dimensionality.

- [ ] The best-fit line
- [ ] The set of curved lines.
- [x] The set of straight lines.

> **Correct**
> Correct, by default linear regression in two dimensions considers all possible straight lines.

- [ ] All hypothesis that give numbers instead of classes.

2. Why do non-linear feature expansions increase model complexity? **1 / 1 point**

- ( ) Because non-linear feature expansions are more complicated than linear features.
- ( ) Because non-linear feature expansions are hard to calculate.
- (•) Because non-linear feature expansions generally increase the size of the hypothesis space.

> **Correct**
> Correct. Increasing the size of the hypothesis space through the addition of non-linear features means more complex hypothesis are available to the learning algorithm.

3. What's the problem with doing regression to find numeric class labels directly? **1 / 1 point**

- ( ) Classification isn't convex.
- ( ) Transfer functions break loss functions.
- ( ) You can't actually convert class labels to numbers.
- ( ) Regression doesn't work for binary values.
- (•) It just works better to separate classes.
- ( ) Classifications are categorical rather than numeric values.

> **Correct**
> True! When you care about class labels, trying to fit a line (or hyperplane) to exactly recreate the classes is more difficult than finding a decision boundary.

4. Why might we *not* want our model to fit perfectly to our training data? **1 / 1 point**

- ( ) The model will have high bias.
- (•) The model will have high variance and not generalize well to new data.
- ( ) We always want our model to fit perfectly to our training data.

Correct

Correct! When our model fits perfectly to our training data we have low bias, but very high variance, resulting in overfitting.

5. What is the definition of a convex function?

1 / 1 point

- (x) For any two points on a graph, the line connecting the points is on or above the line of the graph.
- ( ) A function with neither local minima or global minima.
- ( ) A function with both local minima and global minima.
- ( ) For any two points on a graph, the line connecting the points is on or below the line of the graph.

Correct

Correct! This is exactly the definition of a convex graph.

6. Why do we need iterative functions other than gradient descent to optimize loss functions?

1 / 1 point

- ( ) Because the L2 loss function can have sharp corners.
- ( ) We don't need anything but gradient descent to optimize loss functions.
- (x) Because not all loss functions are differentiable everywhere.

Correct

Correct! In order to apply gradient descent, our loss function must be differentiable everywhere.

7. The bias/variance tradeoff is impacted by (select all that apply)

1 / 1 point

- [x] Model complexity

Correct

Correct! A less complex model means higher bias and lower variance. A more complex model means lower bias and higher variance.

- [x] An overly simple hypothesis space

Correct

Correct! We may have high bias if we choose an overly simply hypothesis space.

- [x] Randomness in training data

Correct

Correct! Each time you sample new data from the same source and run a new analysis, creating a new model.

Most phenomena are noisy, so the data will be different each time due to randomness.

- [ ] Bad dart throwers

8. L1 and L2 regularizers penalize:

1 / 1 point

- (x) The magnitude of weights in the loss function
- ( ) The distance between the line and the training data.
- ( ) The lambda parameter
- ( ) The magnitude of training data

Correct

Correct! The magnitude of weights is a good proxy to measure complexity. Because we don't want our function to be too complex, we use a regularizer.