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.

- [x] The set of straight lines.

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

- [ ] The best-fit line
- [ ] The set of curved lines.
- [ ] All hypothesis that give numbers instead of classes.

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

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

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**

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

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**

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

- [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.

- [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.

- [ ] Bad dart throwers

8. L1 and L2 regularizers penalize: **1 / 1 point**

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

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.