Course › Week 4... › Compr... › Quiz 4

Quiz 4

Problem 1

1/1 point (graded)

A predictor variable is a name for a variable representing which of the following?

● Information that you already know

○ Information that you wish to predict

✓

Submit

Problem 2

1/1 point (graded)

When we fit a line to a set of data, we minimize the mean squared error. Which of the following is the correct equation for the mean squared error?

○ $MSE = \sum_{i=1}^{n}((y^{(i)} - \bar{y})(x^{(i)} - \bar{x}))^2$

○ $MSE = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} + (ax^{(i)} - b))^2$

● $MSE = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - (ax^{(i)} + b))^2$

○ $MSE = \sum_{i=1}^{n}(y^{(i)} - a(x^{(i)} - b))^2$

Processing math: 76%

Submit

Problem 3

1/1 point (graded)

Given the line $y = -3x + 15$, and the points $a = (3, 0)$ and $b = (7, 0)$, which point has the smallest squared error from the line?

○ point a

○ point b

● both have the same squared error

✓

Submit

Problem 4

1/1 point (graded)

In the lecture, we rewrote the loss function, $f(x) = w_1x_1 + w_2x_2 + \ldots + w_dx_d + b$, as a matrix product, $f(x) = \tilde{w} \cdot \tilde{x}$. How did we get $\tilde{w}$?

○ Inserted a 1 at the beginning of the $\mathbf{w}$ vector

○ Inserted a 0 at the beginning of the $\mathbf{w}$ vector

● Inserted the value b at the beginning of the $\mathbf{w}$ vector

✓

Submit

Processing math: 76%

1/1 point (graded)

In order to write the loss function $L(\tilde{w}) = \sum_{i=1}^{n}(y^{(i)} - \tilde{w} \cdot \tilde{x}^{(i)})^2$ in the form *[Math Processing Error]*, we must create a matrix X. If there are n d-dimensional data points, what is the dimension of the matrix X?

- ○ $X \in \mathbb{R}^{n \times d}$
- ● $X \in \mathbb{R}^{n \times (d+1)}$
- ○ $X \in \mathbb{R}^{d \times n}$
- ○ $X \in \mathbb{R}^{(d+1) \times n}$

✓

Submit

Problem 6

1/1 point (graded)

What is the vector $\tilde{w}$ such that the loss function *[Math Processing Error]* is minimized?

- ● $\tilde{w} = (X^T X)^{-1}(X^T y)$
- ○ $\tilde{w} = (X^T X)^{-1}(Xy)$
- ○ $\tilde{w} = X^{-1}(X^T y)$
- ○ $\tilde{w} = (X^T y)(XX^T)^{-1}$

✓

Submit

Problem 7

Processing math: 76%

1/1 point (graded)

What regularizer term does ridge regression use along with the least-squares loss function?

○ *[Math Processing Error]*, where *[Math Processing Error]* is the L_2 norm of w

● *[Math Processing Error]*, where *[Math Processing Error]* is the squared L_2 norm of w

○ *[Math Processing Error]*, where *[Math Processing Error]* is the L_1 norm of w

○ *[Math Processing Error]*, where *[Math Processing Error]* is the squared L_1 norm of w

✓

Submit

Problem 8

1/1 point (graded)

A larger λ in the regularization term for ridge regression will typically result in which of the following?

☐ a larger w

☑ a larger error on the training set

☑ a smaller w

☐ a smaller error on the test set

✓

Submit

Problem 9

Processing math: 76%

Doing linear regression with the Lasso typically results in few features being included in the model.

● True

○ False

✓

Submit

Problem 10

1/1 point (graded)

Suppose our logistic regression model has decision boundary $x_1 + x_2 - 3 = 0$. How would we classify point $p = (1, 3)$?

☑ p is classified as 1 with $> 50\%$ probability

☐ p is classified as 1 with 50% probability

☐ p is classified as 1 with $< 50\%$ probability

✓

Submit

Problem 11

1/1 point (graded)

If you are classifying d-dimensional data using the general linear function $\mathbf{w} \cdot \mathbf{x} + b = 0$ as the probability decision boundary, how would a point x be classified if $\mathbf{w} \cdot \mathbf{x} + b = 2$?

○ a ‘1’ with 12\% probability

○ a ‘1’ with 42\% probability

Processing math: 76%

○ a ‘1’ with 65\% probability

● a ‘1’ with 88\% probability

✓

Submit

Problem 12

1/1 point (graded)
With logistic regression, what value are we trying to optimize?

● The overall probability of the labels of the data points

○ The mean squared error

○ The gradient for the \bf w vector

○ The joint probability distribution between x and y

✓

Submit

Problem 13

1/1 point (graded)
True or False: In logistic regression, the optimal value for \bf w is found by taking the derivative of the loss function, setting it equal to zero, and solving for \bf w.

○ True

● False

✓

Processing math: 76%

Submit

Problem 14

1/1 point (graded)

What does gradient descent do, for a general loss function over a parameter \bf w?

- ○ It finds the exact \bf w needed to minimize the function
- ○ It finds values of \bf w for which the loss function is zero
- ● It finds values of \bf w that approximate local minima of the function
- ○ It provides a closed form solution to \bf w that optimizes the loss function

✓

Submit

Problem 15

1/1 point (graded)

Let’s say we are building a document classifier that will determine if a text is fiction or nonfiction. We decide to use a bag-of-words representation of documents, based on a vocabulary consisting of the 3,000 most commonly used words from text in the training set.

Assume the word “pilot” is found in the test set text but it isn’t one of the 3,000 most commonly found words in the training set. How is the word used in the model?

- ● There is no entry for this word in the vector representation of any document. The word has no impact on the classification.
- ○ The vector representation for that test document has a 0 entry for that word.
- ○ The vector representation for that test document has a 1 entry for that word.

Processing math: 76%

✓

Submit

Problem 16

1/1 point (graded)

True or false: Coefficients in the \bf w vector tend to have a greater impact on the classification of new data as they grow larger.

● True

○ False

✓

Submit

Processing math: 76%