Quiz Submissions - Quiz 2 - Attempt 2

Nhat Le (username: hung.le@mail.mcgill.ca)

Attempt 1

Written: Jan 31, 2019 12:15 AM - Jan 31, 2019 12:35 AM

Submission View

Released: Jan 23, 2019 11:59 PM

View the quiz answers.

Question 1 **0 / 1 point**

Suppose you want to train a QDA model on a classification dataset with 4 classes and 20 different features. How many parameters are necessary to learn for QDA on this dataset?

- 80
- 3360
- ✖ 1600
- ➡ 1680

Hide Feedback

For QDA, you need to learn the mean vectors for each class (4*20=80 parameters), as well as the covariance matrices for each class (4*20*20=1600 parameters), giving 1680 parameters in total.

Question 2 **1 / 1 point**

A company is developing a clinical diagnostic assistant to help uncover a rare abnormality in chest x-rays. The goal of this system is to flag potentially anomalous x-rays for future review. All of the flagged x-rays will be reviewed, but any abnormalities in x-rays that are not flagged by the system may be completely missed

Which of the following statements is most appropriate to this situation:

- Precision is more important than specificity for this system.
- Recall is more important than sensitivity for this system.
- ✔ Sensitivity is more important than precision for this system
- Specificity is more important the sensitivity for this system.

Hide Feedback

In this case, we want the system to have high sensitivity/recall (which are equivalent) because we want to minimize the number of false negatives.

Question 3 **1 / 1 point**

A data scientist has designed a new regularizer called a “logarithmic L2 regularizer” that adds the following penalty to a loss function, based on the weight vector **w**:

$$\mathrm{Err}_{\mathrm{reg}}(\mathbf{w}) = \mathrm{Err}(\mathbf{w}) + \lambda \log_e(m+e) \,||\, \mathbf{w}||_2^2$$

where Err(**w**) denotes the unregularized error for a model, *m* denotes the number of features in the dataset, and *e* denotes Euler's number. Essentially, we have multiplied the standard L2 regularization term by a logarithmic term, based on *m*.

Now, suppose we run the logarithmic L2 regularizer and a standard L2 regularizer with standard gradient descent for linear regression on the exact same dataset and using the exact same value for

$$\lambda$$

.

Which of the following statements is most appropriate:

- ✔ In terms of the bias-variance tradeoff, we would expect the logarithmic L2 regularizer to result in a higher bias model.
- It is impossible to make a general statement.
- In terms of the bias-variance tradeoff, we would expect the logarithmic L2 regularizer to result in a higher variance models.
- The two approaches will result in identical solutions.

Hide Feedback

We know that

$$\log_e(m+e) > 1, \forall m > 1$$

so the "logarithmic" term effectively increases the strength of the regularization, leading to a higher bias and lower variance model.

Question 4 **1 / 1 point**

True or False: In terms of the statistical bias-variance tradeoff, an extremely high bias model is equivalent to a model that is suffering from underfitting.

- ✔ True
- False

Hide Feedback

As discussed in lecture, a model that has very high bias in the bias-variance tradeoff is equivalent to underfitting.

Question 5 **1 / 1 point**

An employee at a movie production company is prototyping a Naive Bayes model to predict whether a movie will be successful (a binary classification task). So far in the prototype there are three binary features:

- *fresh*, which is 1 if the movie is "certified fresh" on Rotten Tomatoes and 0 otherwise.
- *summer,* which is 1 if the movie was released in the summer and 0 otherwise.
- *rock, w*hich is 1 if the movie is starring Dwayne "The Rock" Johnson and 0 otherwise.

Suppose the model is trained *with Laplace add-one smoothing* on the following data:

- *success=1, [fresh=0, summer=0, rock=0]*
- *success=1, [fresh=0, summer=0, rock=0]*
- *success=1, [fresh=1, summer=1, rock=1]*
- *success=0, [fresh=0, summer=1, rock=0]*
- *success=0, [fresh=1, summer=0, rock=0]*

Would this model predict success or failure for a movie with the following attributes: *[fresh=0, summer=0, rock=1]*

- ✔ Success
- Failure
- Impossible to tell (i.e., not enough information given)

Hide Feedback

Using the notation from lecture, the maximum likelihood parameters for this model are:

$$\theta_1 = 3/5, \theta_{1,fresh} = 2/5, \theta_{1,summer} = 2/5, \theta_{1,rock} = 2/5, \theta_{0,fresh} = 1/2, \theta_{0,summer} = 1/2, \theta_{0,rock} = 1/4$$

And from these we can get that

$$P(success = 1|[fresh = 0, summer = 0, rock = 1]) \propto \theta_1(1 - \theta_{1,fresh})(1 - \theta_{1,summer})\theta_{1,rock} = 0.0864$$

and that

$$P(success = 0|[fresh = 0, summer = 0, rock = 1]) \propto \theta_0(1 - \theta_{0,fresh})(1 - \theta_{0,summer})\theta_{0,rock} = 0.025$$

Attempt Score: 4 / 5 - 80 %

Overall Grade (highest attempt): 4 / 5 - 80 %

Done