Quiz Submissions - Quiz 2 - Attempt 1

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

Attempt 1

Written: Jan 30, 2019 10:59 PM - Jan 30, 2019 11:58 PM

Submission View

Released: Jan 23, 2019 11:59 PM

View the quiz answers.

Question 1 **0 / 1 point**

LDA, QDA, and Gaussian Naive Bayes are three different generative approaches to classification. They differ in their statistical assumptions as well as the number of learnable parameters they contain. From the set of options below, please select the correct ordering of the models, in terms of their number of learned parameters (from lowest to highest). Assume that all models are trained on a dataset with 10 classes and 5 input features.

- LDA, Gaussian Naive Bayes, QDA
- QDA, LDA, Gaussian Naive Bayes
- LDA, QDA, Gaussian Naive Bayes
- Gaussian Naive Bayes, LDA, QDA

Hide Feedback

All three models assume that the conditional distribution of the features given the class is given by a multivariate Gaussian distribution. All three must compute the mean vectors for these distributions, so there is no difference there. However, there is a difference in how they compute the covariance matrices:

For Gaussian Naive Bayes, the covariance matrices are assumed to be distinct for each class but diagonal, so that requires 10*5=50 parameters.

For LDA, the covariance matrices are assumed to be the same for the 10 classes, so that requires 5*5=25 parameters.

For QDA, the covariance matrices are assumed to be distinct for the 10 classes, so that requires 10*5*5=250 parameters.

Question 2 **1 / 1 point**

A company developing self-driving cars is working on two systems to detect pedestrians:

System A is used to provide warning to the drivers, and the system designers feel that it is acceptable if System A sometimes unnecessarily warns the driver when no pedestrians are present. However, the system designers feel that it is unacceptable if this System A ever fails to warn the driver when a pedestrian is truly present.

System B is used for automatic braking, and the designers only ever want this system to activate if it is absolutely confident about the presence of a pedestrian.

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

- System A should have high specificity and System B should have high recall.
- System A should have high sensitivity and System B should have high recall.
- System A should have high precision and System B should have high recall.
- ✓ System A should be high recall and System B should be high precision.

Question 3 **1 / 1 point**

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

$$\mathrm{Err}_{\mathrm{reg}}(\mathbf{w}) = \mathrm{Err}(\mathbf{w}) + \sum_{j=1}^{m} w_j^4$$

where Err(**w**) denotes the unregularized error for a model and

$$w_j$$

denotes the j'th entry in the parameter vector **w.**

Which of the following would correspond to the gradient of the error for logistic regression with quartic regularization?

- $$\sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{x}_i\sigma(\mathbf{w}^\top \mathbf{x}_i)) + 4w_i^3$$
- ✓ $$\left[\sum_{i=1}^{n} \mathbf{x}_i(y_i - \sigma(\mathbf{w}^\top \mathbf{x}_i))\right] + \sum_{j=1}^{m} 4w_j^3$$
- $$\sum_{j=1}^{m} w_j^3 + \sum_{i=1}^{n} \mathbf{x}_i(y_i - \sigma(\mathbf{w}^\top \mathbf{x}_i))$$
- $$\left[\sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{x}_i\sigma(\mathbf{w}^\top \mathbf{x}_i))\right] + \sum_{j=1}^{m} 4w_j^3$$

Hide Feedback

To obtain the error of the gradient, we need to take the usual gradient of the error for logistic (see, e.g., Lecture 4 slides 37) and add the gradient of the quartic regularizing term:

$$\nabla_{\mathbf{w}}\left(\sum_{j=1}^{m} w_j^4\right) = \sum_{j=1}^{m} 4w_j^3$$

Question 4 **1 / 1 point**

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

- True
- ✓ False

Hide Feedback

A model that has high-variance in the bias-variance tradeoff is equivalent to overfitting. A model that has very high bias could be 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 on the following data *(without Laplace smoothing):*

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

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

And from these we can get that

$$P(success = 1|[fresh = 0, summer = 0, rock = 1]) \propto \theta_1(1 - \theta_{1,fresh})(1 - \theta_{1,summer})\theta_{1,rock} \approx 0.133$$

and that

$$P(success = 0|[fresh = 0, summer = 0, rock = 1]) \propto \theta_0(1 - \theta_{0,fresh})(1 - \theta_{0,summer})\theta_{0,rock} = 0.05$$

Attempt Score: 4 / 5 - 80 %

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

Done