CSPB 3202 - Truong - Artificial Intelligence

Started on	Saturday, 27 July 2024, 9:36 PM
State	Finished
Completed on	Saturday, 27 July 2024, 9:44 PM
Time taken	8 mins 48 secs
Marks	31.00/31.00
Grade	**15.00** out of 15.00 (**100**%)

Question **1**

Correct

Mark 4.00 out of 4.00

Suppose we perform K-Folds cross-validation on a data set with 100 labeled examples with $K = 5$ and then again with $K = 10$.

When training the classifier on a single fold, how many examples are in the training set in the case that $K = 5$?

80 ✔

When training the classifier on a single fold, how many examples are in the validation set in the case that $K = 10$?

10 ✔

Which value of K above is likely to give the best estimate of the validation error?: 10 ✔

Which value of K above will be the fastest to train?: 5 ✔

Question **2**

Correct

Mark 3.00 out of 3.00

Suppose you train a logistic regression model using the bag-of-words approach to determine if an email is SPAM ($y = 1$) or HAM ($y = 0$) and obtain the following values for the weights:

bias	"money"	"mom"	"work"	"nigeria"
0.5	1.0	-2.0	-1.0	2.0

Note: In the table, the **bias** feature indicates the intercept in the logistic regression model. It does not represent the word "**bias**".

Which word in the logistic regression model is the strongest predictor for HAM? mom ✔

How would the logistic regression model classify an empty email? SPAM ✔

Which of the following emails is **most** likely to be SPAM? {nigeria, work} ✔

Question **3**

Correct

Mark 4.00 out of 4.00

Choose all methods that can help reducing overfitting.

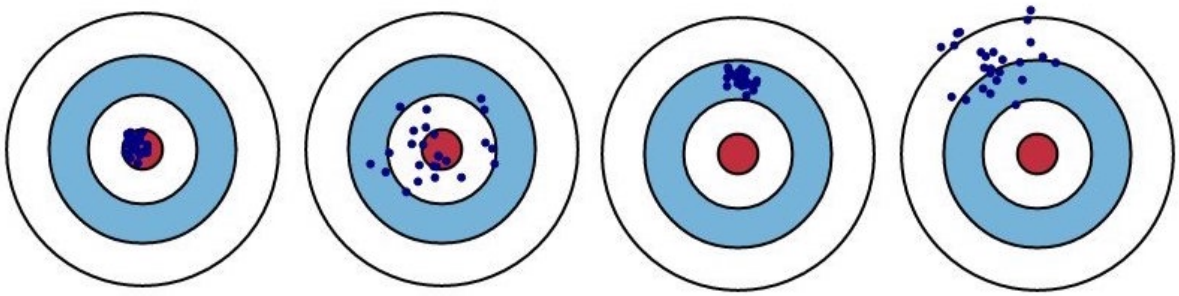
- [x] Increase λ of ridge regression ✔
- [ ] Increase the learning rate of boosting
- [ ] Increase max depth of the tree
- [x] Reduce the number of features ✔
- [x] Add lasso to the loss function ✔
- [ ] Increase the model's variance
- [x] Set aside some part of the train dataset and use it as validation ✔
- [ ] Normalize each feature column
- [x] Cross validate ✔
- [x] Decrease the number of trees in a boosting model ✔
- [x] Increase training data ✔
- [ ] Set ccp_alpha value to 0
- [x] Increase k in KNN model ✔
- [ ] Reduce the C parameter value in SVR
- [ ] Optimize hyperparamters

Question **4**

Correct

Mark 4.00 out of 4.00

We often make an analogy between the bias and variance of a model with target shooting. Consider the four targets shown below. Characterize each model in terms of low/high bias and low/high variance.



How would you characterize first model from the left? low bias and low variance ✔

How would you characterize second model from the left? low bias and high variance ✔

How would you characterize third model from the left? high bias and low variance ✔

How would you characterize fourth model from the left? high bias and high variance ✔

Question **5**

Correct

Mark 3.00 out of 3.00

We are considering launching a new product and wish to know whether it will be a *success* or a *failure*. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.

Fill in the following:

This is a classification ✔ problem involving n= 20 ✔ samples and we are most interested in prediction ✔ .

Question **6**

Correct

Mark 3.00 out of 3.00

Given the following regression model information for the model **SepalLength ~ PetalLength:**

```
In [46]: iris_mdl = smf.ols('SepalLength ~ PetalLength',data=irisdf).fit()
         iris_mdl.summary()
```

Out[46]:

OLS Regression Results

Dep. Variable:	SepalLength	**R-squared:**	0.760
Model:	OLS	**Adj. R-squared:**	0.758
Method:	Least Squares	**F-statistic:**	468.6
Date:	Thu, 19 Apr 2018	**Prob (F-statistic):**	1.04e-47
Time:	15:00:49	**Log-Likelihood:**	-77.020
No. Observations:	150	**AIC:**	158.0
Df Residuals:	148	**BIC:**	164.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.3056	0.078	54.895	0.000	4.151	4.461
PetalLength	0.4091	0.019	21.646	0.000	0.372	0.446

Omnibus:	0.212	**Durbin-Watson:**	1.868
Prob(Omnibus):	0.899	**Jarque-Bera (JB):**	0.350
Skew:	0.070	**Prob(JB):**	0.839
Kurtosis:	2.809	**Cond. No.**	10.3

Complete the following sentence:

The SepalLength ✔ increases by 0.4091 ✔ when the PetalLength ✔ increases by 1.

Question **7**

Correct

Mark 1.00 out of 1.00

Which of the following evaluation metrics is NOT appropriate for measuring the performance of logistic regression on a dataset where the ratio between positive instances and negative instances is 1 : 99?

Select one:

- a. AUC score
- b. Accuracy ✔
- c. Visual comparisons of ROC curves
- d. F1 score

Question **8**

Correct

Mark 1.00 out of 1.00

Assume you have data where the number of predictors **p** is extremely large and the number of observations **n** is small. Would you generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method?

Select one:

- a. Better than inflexible method.
- b. Worse than an inflexible method. ✔

Question **9**

Correct

Mark 1.00 out of 1.00

Assume you have data where the variance of the error terms, *i.e.* $\sigma^2 = Var[\epsilon]$, is extremely high. Would you generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method?

Select one:

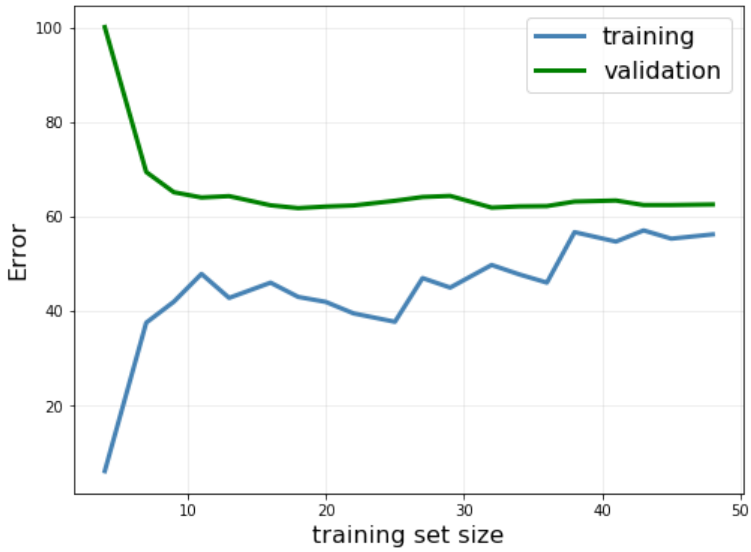
- a. Better than inflexible method.
- b. Worse than an inflexible method. ✔

Question **10**

Correct

Mark 2.00 out of 2.00

Suppose that we are building a model for a particular data set and we have reason to believe that we can achieve a training error around 10 with the right choice of hyperparameters. Suppose also that you've fit an initial model which produces the following learning curve. How would you characterize this model?



Select one:

- a. Low Bias / High Variance
- b. Low Bias / Low Variance
- c. High Bias / High Variance
- d. High Bias / Low Variance ✔

Question **11**

Correct

Mark 1.00 out of 1.00

Suppose that you are performing a binary classification to assign a class label $y \in \{0, 1\}$ to each data point and you model the probability that data point x belongs to Class 1 by $p(y = 1 \mid x) = \text{sigm}(\hat{\beta_0} + \hat{\beta_1} x)$ where $\hat{\beta_0} = 1$ and $\hat{\beta_1} = 1$. How would your model classify a data point with $x = -1.5$.

Select one:

- a. It would predict 0.5
- b. Class 1
- c. Inconclusive
- d. Class 0 ✔

Question **12**

Correct

Mark 1.00 out of 1.00

Assume you are modeling $Y \sim X$ where X is a discrete or categorical variable and Y is a continuous variable.

Is this a regression or classification problem?

Select one:

- a. Either technique could be used.
- b. Regression ✔
- c. Classification

Question **13**

Correct

Mark 3.00 out of 3.00

Given the following confusion matrix:

A	**49**	0	**0**
B	**1**	20	13
C	**0**	16	**35**
Yt/Yp	**A**	**B**	**C**

What is the classifier *recall* for class **B**? Enter your response as a fraction? The answer should have the form of an irreducible fraction n/m, where n and m are integer numbers (irreducible means n and m do not have common divisor other than 1).

10/17

Your last answer was interpreted as follows: $\frac{10}{17}$