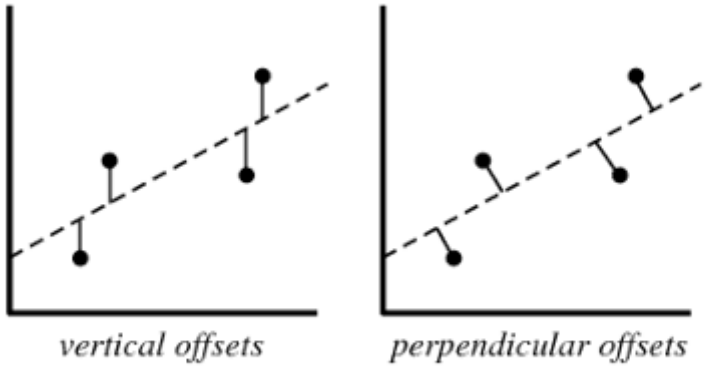
Week 4 Quiz

Total points 15/20

Graded Quiz for Week 4

15 of 20 points

✓ Which of the following offsets, do we use in linear regression's least square line fit? Suppose horizontal axis is independent variable and vertical axis is dependent variable. * 1/1



- None of above
- Both, depending on the situation
- Vertical offset ✓
- Perpendicular offset

✓ Which of the following evaluation metrics can not be applied in case of logistic regression output to compare with target? 1/1

- ○ AUC-ROC
- ○ Accuracy
- ○ Logloss
- ◉ Mean-Squared-Error ✓

✓ Suppose, you got a situation where you find that your linear regression model is under fitting the data. In such situation which of the following options would you consider? 1/1

- ☑ I will add more features ✓
- ☑ I will start introducing polynomial degree variables ✓
- ☐ I will remove some variables

✗ Now situation is same as written in previous question(under fitting).Which of following regularization algorithm would you prefer? 0/1

- ◉ A) L1 ✗
- ○ B) L2
- ○ C) Any
- ○ D) None of these

✓ Suppose that you have a dataset D1 and you design a linear regression model of degree 3 polynomial and you found that the training and testing error is “0” or in another terms it perfectly fits the data. What will happen when you fit degree 4 polynomial in linear regression? 1/1

- (●) A) There are high chances that degree 4 polynomial will over fit the data ✓
- ( ) B) There are high chances that degree 4 polynomial will under fit the data
- ( ) C) Can’t say
- ( ) D) None of these

✓ Continuing the previous question, What will happen when you fit degree 2 polynomial in linear regression? 1/1

- ( ) A) It is high chances that degree 2 polynomial will over fit the data
- (●) B) It is high chances that degree 2 polynomial will under fit the data ✓
- ( ) C) Can’t say
- ( ) D) None of these

✓ If we are building a system to predict if a person has cancer or not, what metric is better? 1/1

- (●) Recall ✓
- ( ) Precision

✓ If we are building a system to predict if we should decrease the credit limit on a particular account, which metric is better? 1/1

- (●) Precision ✓
- ( ) Recall

✓ _____ are the cases where you wrongly classified a non-event as an event a.k.a Type I error. 1/1

- (●) False Positives ✓
- ( ) False Negatives
- ( ) True Positives
- ( ) Truen Negatives

✓ The effectiveness of an SVM depends upon: 1/1

- ( ) A) Selection of Kernel
- ( ) B) Kernel Parameters
- ( ) C) Soft Margin Parameter C
- (●) D) All of the above ✓

✓ If I am using all features of my dataset and I achieve 100% accuracy on my training set, but ~70% on validation set, what should I look out for? 1/1

- ○ A) Underfitting
- ○ B) Nothing, the model is perfect
- ◉ C) Overfitting ✓

✓ Which of the following is/are true about bagging trees? 1/1

- ○ In bagging trees, individual trees are independent of each other
- ○ Bagging is the method for improving the performance by aggregating the results of weak learners
- ◉ Both the above ✓
- ○ None of the above

✓ Which of the following is/are true about boosting trees? 1/1

- ○ In boosting trees, individual weak learners are independent of each other
- ◉ It is the method for improving the performance by aggregating the results of weak learners ✓
- ○ Both the above
- ○ None of the above

✓ Which of the following is true about “max_depth” hyperparameter in Gradient Boosting? 1/1

- [x] Lower is better parameter in case of same validation accuracy ✓
- [ ] Higher is better parameter in case of same validation accuracy
- [x] Increase the value of max_depth may overfit the data ✓
- [ ] Increase the value of max_depth may underfit the data

✓ Which of the following statements are true? Check all that apply. 2/2

- [x] A model with more parameters is more prone to overfitting and typically has higher variance. ✓
- [ ] If the training and test errors are about the same, adding more features will not help improve the results.
- [x] If a learning algorithm is suffering from high bias, only adding more training examples may not improve the test error significantly. ✓
- [x] If a learning algorithm is suffering from high variance, adding more training examples is likely to improve the test error. ✓
- [x] When debugging learning algorithms, it is useful to plot a learning curve to understand if there is a high bias or high variance problem. ✓

✗ Suppose a massive dataset is available for training a learning algorithm. Training on a lot of data is likely to give good performance when two of the following conditions hold true.Which are the two? 0/2

- [ ] We train a learning algorithm with a large number of parameters (that is able to learn/represent fairly complex functions).
- [x] The features x contain sufficient information to predict accurately. (For example, one way to verify this is if a human expert on the domain can confidently predict when given only). ✓
- [x] We train a learning algorithm with a small number of parameters (that is thus unlikely to overfit). ✗
- [ ] We train a model that does not use regularization.

✗ Suppose you are working on a spam classifier, where spam emails are positive examples (y = 1) and non-spam emails are negative examples (y = 0). You have a training set of emails in which 99% of the emails are non-spam and the other 1% is spam.Which of the following statements are true? Check all that apply. 0/2

- [x] A good classifier should have both a high precision and high recall on the cross validation set. ✓
- [ ] If you always predict non-spam (output y=0), your classifier will have an accuracy of 99%.
- [x] If you always predict non-spam (output y=0), your classifier will have 99% accuracy on the training set, but it will do much worse on the cross validation set because it has overfit the training data. ✗
- [ ] If you always predict non-spam (output y=0), your classifier will have 99% accuracy on the training set, and it will likely perform similarly on the cross validation set.

BEFORE SUBMITTING

The next page will lead you to accept the honor code and submit the User ID and Password. Make sure you have these ready.

THE USER ID AND PASSWORD FOR ALL QUIZZES/ASSIGNMENTS IS SAME THROUGHOUT COURSE

User Details 0 of 0 points

Honor Code

1) You can submit the form only once, using a single User ID and Pass. Usage of multiple accounts for submission of quizzes will lead to harsh actions being taken.

2) Your answers to the questions must be your own work. (Very Important)

3) You may not share your solutions with anyone else unless explicitly permitted by the mentor. This includes anything written by you, as well as any official solutions provided by the course .

4) You may not engage in any other activities that will dishonestly improve your results or dishonestly improve or damage the results of others.

You can report Honor Code violations by contacting any of the members of Consulting and Analytics Club, IIT Guwahati.

Name *

Dhrubanka Dutta

Email ID *

dhrubanka.99dutta@gmail.com

User ID *

U011W5850DA

Password *

GN7Y34JB