DATA ANALYTICS WITH PYTHON

MIDTERM EXAM

MM: 30

Name:

ID:

17th April 2019

Q1. Write whether true or false with one-line justification. (Tread very carefully) - 5M

a. The time it takes to make a prediction using a decision tree depends on the number of nodes in that tree

False, depends on number of levels

b. In Ridge Regression, no matter how large a regularization constant $\lambda > 0$ we set, we will always get good solutions

False, at high λ algorithm tend to penalize even the good solutions

c. You need a classifier to be deployed for an end-to-end application with tight latency-constraint. Do you reject kNN straight away. Why?

True, kNN is instance based it computes every time for a new input making it slow

d. Gradient-based methods for loss optimization always converge to global minima

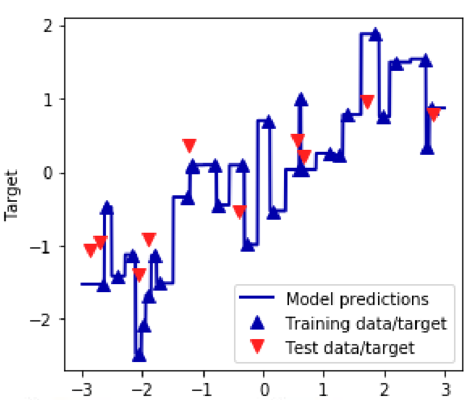
False, they may converge to local minima as well

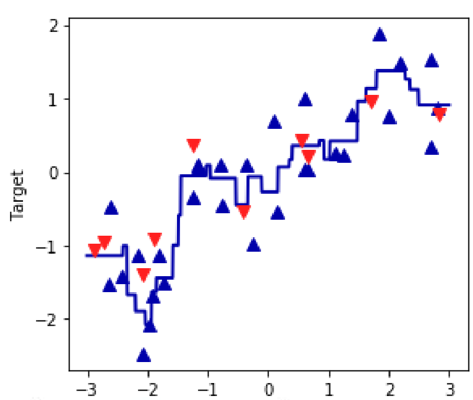
e. Soft-Margin SVM is better at avoiding overfitting

True, it allows maximum margin with some misclassification

Q2. Which of these pictures represent larger value of k for k-NN regression? Which is better? Justify. - 2M

Ans: Second Picture represents larger value of k. It is better since it avoids overfitting.





Q3. Which of the following could improve the model? Why do you think so? - 2M

Traning AUC	Test AUC	Desired Test AUC
0.95	0.75	0.90

a. **Add more training data** ⟵ **Right** **Sincenwe have enough features to get 95% training accuracy but not enough variability in our test set to capture general trend. More training data is expected to fill in for that**

b. Add more features

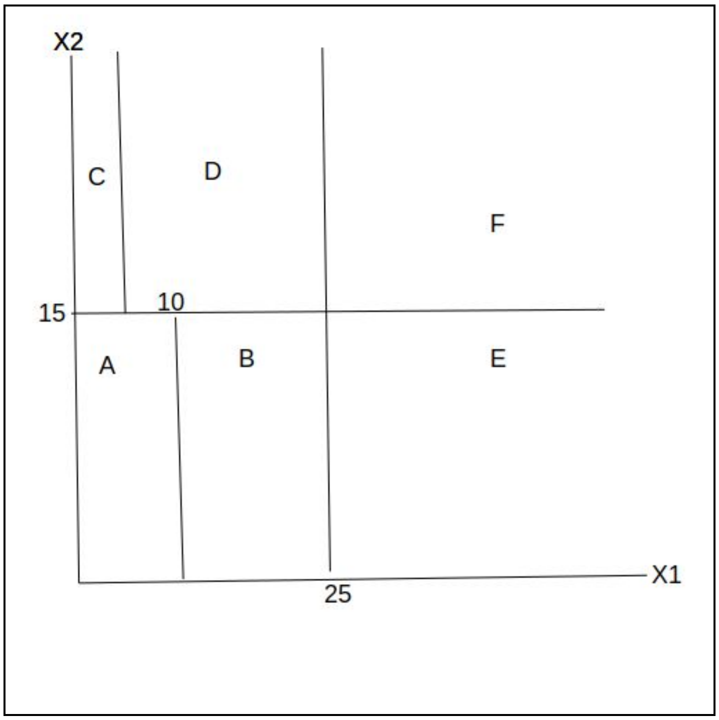
Q4. Explain how is Logistic Regression different from Support Vector Machines. - 3M

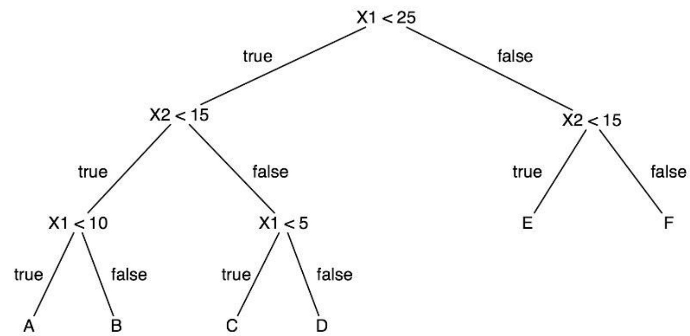
Ans: 1. SVM try to maximize the margin between the closest support vectors while LR the class probability.

2. LR is more sensitive to outliers than SVM because the cost function of LR diverges faster than those of SVM.

3. Logistic Regression produces probabilistic values while SVM produces 1 or 0.

Q5. Draw the decision boundary of given decision tree. - 3M





Q6. Why do large margins in SVM mean good generalization? - 2M

Ans: Large SVM margins correspond to small L2 norm which corresponds to regularized or simple solutions that is the coefficients are small which inherently controls overfitting and provide good generalization.

Q7. Write the loss function for Logistic Regression as discussed in the class. Also explain how can we solve it for *w*? - 3M

Ans:

Either of these

$$L(\boldsymbol{w}) = \sum_{n=1}^{N} \ell(y_n, f(\boldsymbol{x}_n)) = \sum_{n=1}^{N} [-y_n \log(\mu_n) - (1 - y_n)\log(1 - \mu_n)]$$

$$\ell(y_n, f(\boldsymbol{x}_n)) = \begin{cases} -\log(\mu_n) & y_n = 1 \\ -\log(1 - \mu_n) & y_n = 0 \end{cases}$$

Solving for W : Differentiating loss function and equating to zero → Using Gradient Descent to find a minima

Q8. Write various ways to avoid overfitting in Decision Trees? - 2M

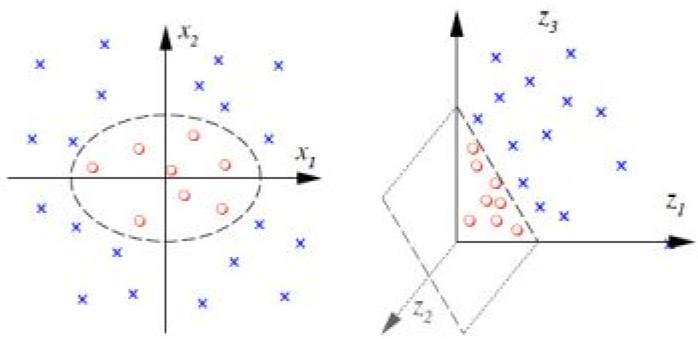
Ans:

Pre-Pruning : Prune while building the tree (stopping early)

Post-Pruning: Prune after building the tree

Q9. Describe with help of an example how Kernel trick gives immense power to SVMs? - 3M

Ans:



Kernels, using a feature mapping φ, map data to a new space where the original learning problem becomes "easy". However directly computing in this feature space is expensive. Kernel function k implicitly defines an associated feature mapping φ.

φ takes input x ∈ X (input space) and maps it to F (new "feature space")

The kernel function k can be seen as taking two points as inputs and computing their inner-product based similarity in the F space.

This computation happens in original feature space hence no extra overhead.

This allows SVMs to classify even non-linear patterns without actually computing non-linearly.

(BONUS QUESTION) - 5M*

Suppose you want to train a perceptron on the following dataset:

X1	X2	Y
2	6	-1
1	3	1
3	9	1

Give a brief intuitive explanation for why the perceptron cannot learn this task. Then give a proof using inequalities expressed in terms of the weights W1, W2, and b.

Hint : Perceptron Classification Equation

Ans : An intuitive explanation + some way to show equations have no solutions.