Foundation of ML Quiz - 2

Your email will be recorded when you submit this form

Foundations of ML

Assume that we have a training set of fixed size N and that all features are uniformly distributed on [0; 1]. Predicted response y correspond to the average of the responses associated to the training examples that are near x. Suppose that we have three features (d = 3) and we want to predict using only training examples that are within 10% of the input range in all the dimensions. On average, what fraction of the training examples will we use to make each prediction? [Marks :2]

- 10%
- 5%
- 1%
- 0.1%

Suppose, you have given the following data where x and y are the 2 input variables and Class is the dependent variable. Suppose, you want to predict the class of new data point x=1 and y=1 using euclidian distance in 3-NN. In which class this data point belongs to? [Marks : 2]

x	y	Class
-1	1	$-$
0	1	$+$
0	2	$-$
1	-1	$-$
1	0	$+$
1	2	$+$
2	2	$-$
2	3	$+$

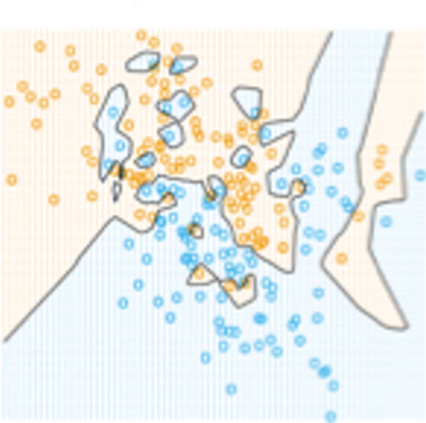
- + class
- - class
- cant say
- Both the class

Consider the data below, and assume you use a decision tree to perform regression on this data. which of the following is correct ? [Marks : 4]

humidity	wind	Golf players
high	Weak	2
high	strong	4
high	weak	1
normal	strong	5
normal	weak	3

- ○ Variance of the of the root node is 1 and first split is based on humidity feature
- ○ Variance of the root node is 1 and first split is based on wind feature
- ○ Variance of root node is 2 and first split is based on humidity feature
- ○ Variance of root node is 2 and first split is based on wind feature
- ○ Variance of root node is 0.33 and first split is based on humidity feature

Which classifier might have generated the following decision boundary [Marks : 2]



- Linear SVM
- Logistic regression
- 1-NN
- quadratic discriminant analysis
- None

Given training data, for test sample (1,1), using k-NN with k=3 choose the correct option. For weighted KNN use the inverse of Euclidean distance as weight. [Marks : 3]

$$\begin{bmatrix} x & y & class \\ 0 & 4 & 0 \\ 0 & 3 & 0 \\ 2 & 1 & 1 \end{bmatrix}$$

- class predicted by K-NN is 1 and by weighted K-NN is 0
- class predicted by K-NN is 0 and by weighted K-NN is 0
- class predicted by K-NN is 0 and by weighted K-NN is 1
- class predicted by K-NN is 1 and by weighted K-NN is 1

Choose the correct option. 1. Bagging can be done in parallel 2. Bagging helps in reducing overfitting. [Marks : 1]

- False, True
- True, True
- True, False
- False,False

You are working on a spam classification system using a binary classifier. “Spam” is a positive class (y = 1) and ”not spam” is the negative class (y = 0). You have trained your classifier and there are m = 1000 examples in the cross-validation set. The chart of predicted class vs. actual class is given below. What is the classifier’s precision? [Marks : 2]

	Actual Class:1	Actual Class:0
Predicted Class: 1	85	890
Predicted Class: 0	15	10

- 0.087
- 0.85
- 0.095
- 0.90
- 1

Which of the following is true about weak learners used in the ensemble model? [Marks : 2]

- High bias, Low Variance
- High bias, High Variance
- Low bias, Low Variance
- Low bias, High Variance

Suppose in a binary classification problem, you have given the following predictions of three models (M1, M2, M3) for five observations of the test data set. When using the weighted voting method, which of the following will be the output of an ensemble model. Hint: Count the vote of M1, M2, and M3 as 2.5 times, 6.5 times, and 3.5 times respectively. [Marks : 3]

M1	M2	M3	Output
1	1	0	
0	1	0	
0	1	1	
1	0	1	
1	1	1	

- [0,1,0,0,1]
- [1,0,1,1,1]
- [1,1,1,0,1]
- [1,1,1,1,1]
- None of the above

[True/False] The error of hypothesis in Adaboost is computed as number of misclassified examples divided by total number of examples. [Marks : 1]

- True
- False

[True/False] Consider the hyper-cube and the inscribed hyper-sphere. The ratio of the volume of the sphere to volume of the cube increases with dimension d. [Marks : 2]

- True
- False

In the Adaboost algorithm, which of the following reweighting schemes are most appropriate ? Assume, yi is actual label (+1/-1) and h(xi) is the predicted label by a weak classifier for data point xi, W(i) is weight associated with it and a is a constant. [Marks : 2]

- W(i) = W(i) exp(-a yi h(xi))
- W(i) = W(i) exp(a yi h(xi))
- W(i) = W(i) log (a yi h(xi))
- W(i) = W(i) log (-a yi h(xi))
- W(i) = W(i) a yi h(xi)

Consider X and Y are discrete random variables; X has 6 possible states (values) and Y has 5 possible states. A general joint distribution on two such variables would require minimum A parameters to define it (considering sum-to-one constraint). Assume $p(x, y) = p(x)p(y)$, where $X \perp Y$ (X independent of Y). By assuming (unconditional) independence, we need minimum B parameters (considering sum-to-one constraint) to define $p(x, y)$. Then, [Marks : 2]

- A = 30, B = 11
- A = 20, B = 9
- A = 29, B = 9
- A = 20, B = 10

Which of the following are most appropriate ? [Marks : 1]

- Quadratic discriminant analysis is a discriminative model
- quadratic discriminant analysis is equivalent to logistic regression
- Quadratic discriminant with diagonal covariance is equivalent to Gaussian Naive Bayes
- Quadratic discriminant analysis is less prone to overfitting

Naive Bayes Classifier : Consider 2 classes HAM and SPAM, and assume bag of words (binary valued features i.e. 1 if a word is present in the document else 0) representation of documents over vocabulary set V = {good, bad, very}. Assuming a multivariate Bernoulli naive Bayes model to do classification, which of the following statements about classifying the document { very good } is correct ? [Marks : 4]

HAM examples

d1: {good}

d2: {very good}

d3 : {bad}

SPAM examples

D4 : {bad}

D5 : {very bad}

D6 : {very bad very bad}

D7 : {good}

- Probability of classifying to SPAM class is zero
- Probability of classifying to SPAM class is 1/7
- Probability of classifying to HAM class is 1/7
- probability of classifying to SPAM class is 2/21
- Probability of classifying to SPAM class is 1/14

You are working on a spam classification system using a binary classifier. “Spam” is a positive class (y = 1) and ”not spam” is the negative class (y = 0). You have trained your classifier and there are m = 1000 examples in the cross-validation set. The chart of predicted class vs. actual class is given below. What is the classifier’s recall ? [Marks : 2]

	Actual Class:1	Actual Class:0
Predicted Class: 1	85	890
Predicted Class: 0	15	10

- 0.087
- 0.85
- 0.095
- 0.90
- 1

The distance of the point (2,3,-5) from the plane x+2y-2z=9 is - [Marks : 2]

- 3
- 2
- 3/2
- 2/3
- 1

Consider the training dataset given below - we'd like to build a decision stump with this dataset. what is the entropy at the root of the tree ? [Marks : 2]

f_1	f_2	Label
4	1	1
6	6	0
9	5	1
1	2	0
7	3	1
5	4	0

- 1
- 0.25
- 0.5
- 0

Given N number of d-dimensional data [Xi] (i=1...N), you run principle component analysis and pick P principle components. Can you always reconstruct any data point [Xi] for i=1..N, from the P principle components with zero reconstruction error? [2 Marks]

- Yes, if P < d
- Yes, if P = d
- Yes, if P < N
- No always

Principal component analysis aims to find a project matrix such that

- projected variance is minimum and distortion is maximum
- projected variance is maximum and distortion is maximum
- projected variance is minimum and distortion is minimum
- projected variance is maximum and distortion is minimum

[True/False] In decision trees, same feature can be selected multiple times during tree construction [Marks : 1]

- False
- True

Let A be a real, symmetric n x n matrix. Which of the following are true about A's eigenvectors and eigenvalues? [1 Mark]

- ◯ A can have no more than n distinct eigenvalues
- ◯ The vector 0 is an eigenvector, because A0 = lambda 0
- ◯ A can have no more than 2n distinct unit-length eigenvectors
- ◯ We can find 2n mutually orthogonal eigenvectors of A

Consider the training dataset given below - we'd like to build a decision stump with this dataset. What is the rule for the first split if using entropy as the splitting condition? [Marks : 4]

f_1	f_2	Label
4	1	1
6	6	0
9	5	1
1	2	0
7	3	1
5	4	0

- f1 >= 5
- f2 >= 3
- f2 >= 5
- f1 >= 7

[True/False] Decision boundaries obtained using quadratic discriminant analysis and logistic regression has same form. [Marks : 1]

- True
- False

Back Submit Page 2 of 2 Clear form