CS461 Quiz One

0. True / False Questions

1. Linear regression models are not capable of learning a non-linear functional relationships. (True / False)

2. When the prior distribution is uniform, MAP becomes equivalent to ML estimation. (True / False)

3. In ridge regression, as the regularization parameter increases, the model complexity increases. (True / False)

4. In maximum likelihood estimation, the variance of an estimator can be reduced by increasing number of data points (True / False)

5. Both Naive Bayes and logistic regression models learn posterior density $P(C_k|x)$. (True / False)

6. Perceptron algorithm converges regardless of whether the data is linearly separable. (True / False)

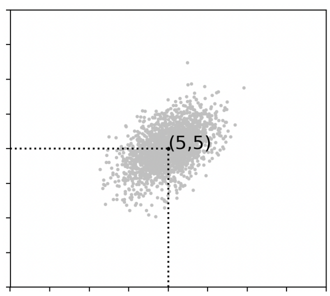
7. Logistic regression converges regardless of whether the data is linearly separable. (True / False)

8. When data is linearly separable, logistic regression's sigmoid function becomes infinitely steep around the decision boundary. (True / False)

9. Even if the data is linearly separable, the perceptron algorithm may not converge depending on the choice of initial points. (True / False)

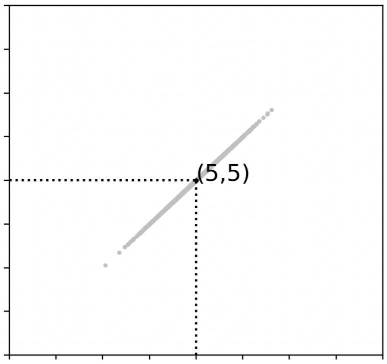
10. Perceptron algorithm is primarily designed for binary classification. (True / False)

2. Whitening data is a prepossessing step in machine learning. Suppose we have 2,000 2-D data points and computed mean and covariance information as below.



$$E[X] = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \quad COV[X,X] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Q: PCA approximation for a data point x is given by $x' = \bar{x} + E_M E_M^t (x - \bar{x})$. When PCA approximation for the 2,000 data points are shown below, what are E_M and $\bar{x}$?

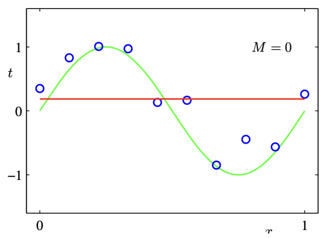


- $E_M(2 \times 1) = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$
- $\bar{x}(2 \times 1) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$

Q: what is be the variance (σ^2) $E_M^t(x - \bar{x})$?

4. The figures below show the two problematic situations : overfitting and underfitting. Please circle the related symptoms and possible solutions for each problem. (the sinusoidal wave represents the ground truth, dots indicate the collected data, and the final curve depicts the model we have learned.

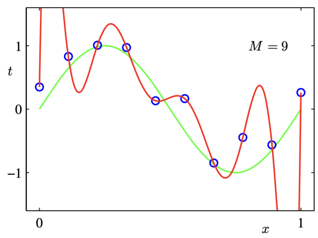
1.1



1.1 symptoms	
high train error	
high test error	
high bias	

1.1 solutions	
collect more data	
regularization	
increase the number of basis functions	

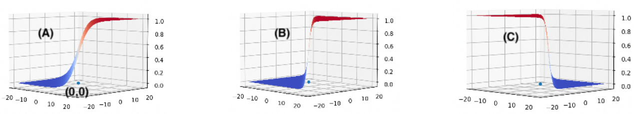
1.2



1.2 symptoms	
low train error	
high test error	
high bias	

1.2 solutions	
collect more data	
regularization	
increase the number of basis functions	

5. Suppose you trained a logistic regression model given three data sets and learned the three sigmoid functions $P(C_1|x)$ as below. Based on their orientation and steepness determine their decision rules.



① $0.3x + 0.3y \underset{C_0}{\overset{C_1}{\lesseqgtr}} 0$

② $x + y \underset{C_0}{\overset{C_1}{\lesseqgtr}} 0$

③ $-x - y \underset{C_0}{\overset{C_1}{\lesseqgtr}} 0$

0. True / False Questions

1. Linear regression models are not capable of learning a non-linear functional relationships. (True / **False**)

2. When the prior distribution is uniform, MAP becomes equivalent to ML estimation. (**True** / False)

3. In ridge regression, as the regularization parameter increases, the model complexity increases. (True / **False**)

4. In maximum likelihood estimation, the variance of an estimator can be reduced by increasing number of data points (**True** / False)

5. Both Naive Bayes and logistic regression models learn posterior density $P(C_k|x)$. (**True** / False)

6. Perceptron algorithm converges regardless of whether the data is linearly separable. (True / **False**)

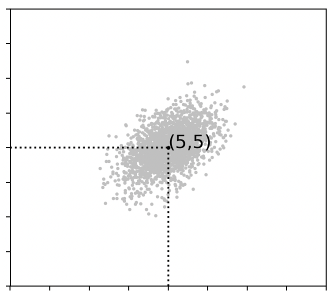
7. Logistic regression converges regardless of whether the data is linearly separable. (**True** / False)

8. When data is linearly separable, logistic regression's sigmoid function becomes infinitely steep around the decision boundary. (**True** / False)

9. Even if the data is linearly separable, the perceptron algorithm may not converge depending on the choice of initial points. (True / **False**)

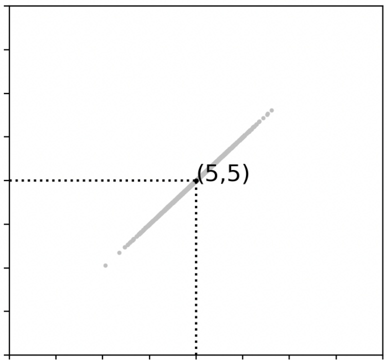
10. Perceptron algorithm is primarily designed for binary classification. (**True** / False)

2. Whitening data is a prepossessing step in machine learning. Suppose we have 2,000 2-D data points and computed mean and covariance information as below.



$$E[X] = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \quad COV[X,X] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Q: PCA approximation for a data point x is given by $x' = \bar{x} + E_M E_M^t (x - \bar{x})$. When PCA approximation for the 2,000 data points are shown below, what are E_M and $\bar{x}$?



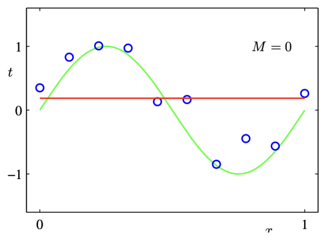
- $E_M(2 \times 1) = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

- $\bar{x}(2 \times 1) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$

Q: what is be the variance (σ^2) $E_M^t(x - \bar{x})$? 3

4. The figures below show the two problematic situations : overfitting and underfitting. Please circle the related symptoms and possible solutions for each problem. (the sinusoidal wave represents the ground truth, dots indicate the collected data, and the final curve depicts the model we have learned.

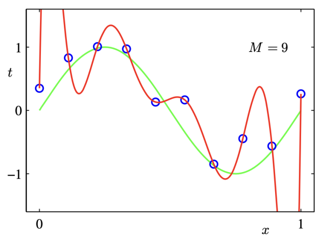
1.1



1.1 symptoms	
high train error	✓
high test error	✓
high bias	✓

1.1 solutions	
collect more data	
regularization	
increase the number of basis functions	✓

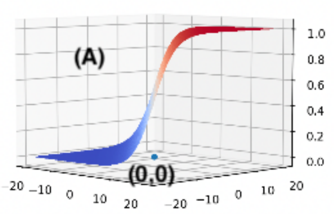
1.2

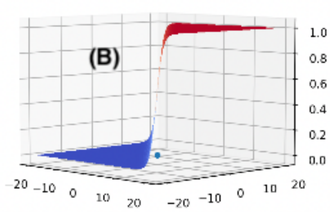


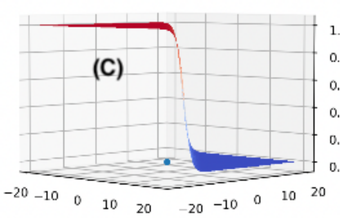
1.2 symptoms	
low train error	✓
high test error	✓
high bias	

1.2 solutions	
collect more data	✓
regularization	✓
increase the number of basis functions	

5. Suppose you trained a logistic regression model given three data sets and learned the three sigmoid functions $P(C_1|x)$ **as below. Based on their orientation and steepness determine their decision rules.**







① $0.3x + 0.3y \underset{C_0}{\overset{C_1}{\lesseqgtr}} 0$ (A)

② $x + y \underset{C_0}{\overset{C_1}{\lesseqgtr}} 0$ (B)

③ $-x - y \underset{C_0}{\overset{C_1}{\lesseqgtr}} 0$ (C)