CS461 Section #:	
Name:	
NetID:	

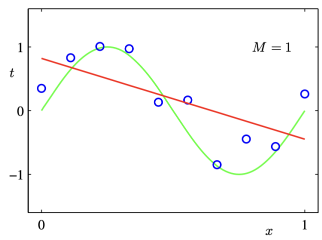
CS461 Quiz One

0. True / False Question .

- Maximum likelihood estimation is sensitive to the size of dataset. (**True** / False)
- Linear regression models are not capable of learning a non-linear functional relationships. (True / **False**)
- When the prior distribution is uniform, MAP becomes equivalent to ML estimation. (**True** / False)
- In ridge regression, as the regularization parameter increases, the model complexity increases. (True / **False**)
- Ridge regression is a special case of MAP when both the prior and likelihood are Gaussian. (**True** / False)

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

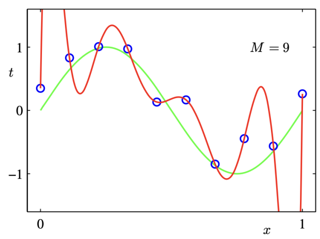
1.1



1.1 symptoms	
high train error	○
high validation error	○
high variance	

1.1 solution	
collect more data	
regularized regression	
increase # of basis functions	○

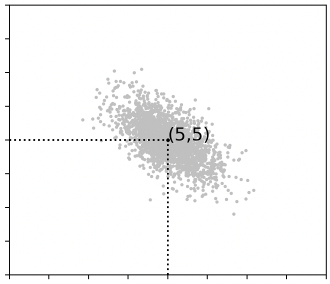
1.2



1.2 symptoms	
high train error	
high validation error	○
high variance	○

1.2 solution	
collect more data	○
regularization regression	○
increase # of basis functions	

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.



$$E[X] = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \quad COV[X, X] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

2.1 Let Y be a random vector defined by $\vec{Y} = A\vec{X} + \vec{b}$. Express $E[Y]$ and $COV[Y, Y]$ in terms of $E[X]$, $COV[X, X]$, A, and b.

$$\begin{aligned} COV(Y, Y) &= E[(Y - E[Y])(Y - E[Y])^t] \\ &= E[(AX + b - AE[X] - b)(AX + b - AE[X] - b)^t] \\ &= AE[(X - E[X])(X - E[X])^t]A^t \\ &= A \cdot COV(X, X) \cdot A^t \end{aligned}$$

$$E[Y] = E[AX + b] = AE[X] + b$$

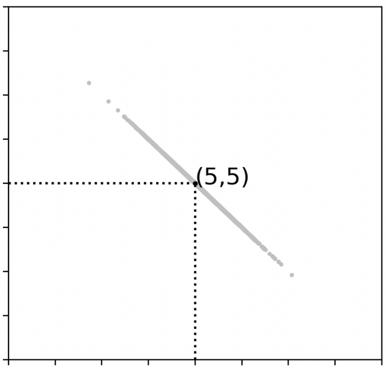
2.2 Design A and b to whiten Y. i.e. $E[Y] = 0$ and $COV[Y, Y] = I$. You don't need to compute the matrix multiplication. Just leave them in block form.

$$A = \Lambda^{-1/2} \cdot U^t = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \tag{1}$$

$$b = -AE[X] = -\begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 5 \end{bmatrix} \tag{2}$$

$$\tag{3}$$

2.3 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 $\bar{x}$?



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

3. Suppose we collected four data points (x_1, y_1), ..., (x_4, y_4) to estimate a linear relationship between x and y, represented by $y = w_0 + w_1 x$.

data num	(x, y)
d_1	(2 , 3)
d_2	(3 , 4)
d_3	(4 , 5)
d_4	(5 , 6)

2.1 Please represent the four data points in the data-matrix Φ, which has the size of 4×2 .

$$\Phi = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}$$

2.1 The normal equation to compute $\vec{W} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$ is shown below. Do you think the equation will give you an MMSE approximated solution $\vec{W}'$ or an exact solution $\vec{W}$?

$$\Phi^t \Phi \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \Phi^t \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

sol) it will be an exact equation because $\vec{y} = [3, 4, 5, 6]$ is on the same hyperplane with $[1, 1, 1, 1]$ and $[2, 3, 4, 5]$.

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \cdot \vec{W} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$