1   K-Nearest Neighbors

1. **Select all that apply:** Please select all that apply about k-NN in the following options:

   ☐ k-NN works great with a small amount of data, but it is too slow when the amount of data becomes large.

   ☐ k-NN is sensitive to outliers; therefore, in general, we decrease k to avoid overfitting.

   ☐ k-NN can only be applied to classification problems, and it cannot be used to solve regression problems.

   ☐ We can always achieve zero training error (perfect classification) on a consistent data set with k-NN, but it may not generalize well in testing.

2. (1 point) **Select one.** A k-Nearest Neighbor model with a large value of k is analogous to...

   ○ A *sho*

   ○ A *sho*

   ○ A *long* Decision Tree with a *low* bra

   ○ A *long* Decision Tree with a *high* br

3. (1 point) **Select one.** Imagine you are using a k-Nearest Neighbor classifier on a data set with lots of noise. You want your classifier to be *less* sensitive to the noise. Which is more likely to help and with what side-effect?

   ○ Increase the value of $k \rightarrow$ Increase in prediction time

   ○ Decrease the value of $k \rightarrow$ Increase in prediction time

   ○ Increase the value of $k \rightarrow$ Decrease in prediction time

   ○ Decrease the value of $k \rightarrow$ Decrease in prediction time

2   Model Selection and Errors

1. **Training Sample Size:**   In this problem, we will consider the effect of training sample size N on a linear regression problem with M features.

   The following plot shows the general trend for how the training and testing error change as we increase the training sample size N. Your task in this question is to analyze this plot and identify which curve corresponds to the training and test error. Specifically:

   1. Which curve represents the training error? **Please provide 1–2 sentences of justification.**

   2. In one word, what does the gap between the two curves represent?

3   Linear Regression

1. (1 point) **Select one:** The closed form solution for linear regression is $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$. Suppose you have $N = 35$ training examples and $M = 5$ features (excluding the bias term). Once the bias term is now included, what are the dimensions of $\boldsymbol{X}$, $\boldsymbol{y}$, $\hat{\boldsymbol{\theta}}$ in the closed form equation?

   ○ $\boldsymbol{X}$ is 35×6, $\boldsymbol{y}$ is 35×1, $\hat{\boldsymbol{\theta}}$ is 6×1

   ○ $\boldsymbol{X}$ is 35×6, $\boldsymbol{y}$ is 35×6, $\hat{\boldsymbol{\theta}}$ is 6×6

   ○ $\boldsymbol{X}$ is 35×5, $\boldsymbol{y}$ is 35×1, $\hat{\boldsymbol{\theta}}$ is 5×1

   ○ $\boldsymbol{X}$ is 35×5, $\boldsymbol{y}$ is 35×5, $\hat{\boldsymbol{\theta}}$ is 5×5

2. Consider linear regression on N 1-dimensional points $x^{(i)} \in \mathbb{R}$ with labels $y^{(i)} \in \mathbb{R}$. We apply linear regression in both directions on this data, i.e., we first fit y with x and get $y = \beta_1 x$ as the fitted line, then we fit x with y and get $x = \beta_2 y$ as the fitted line. Discuss the relations between β_1 and β_2:

   True or False: The two fitted lines are always the same, i.e. we always have $\beta_2 = \frac{1}{\beta_1}$.

   ○ True

   ○ False

3. Please circle **Tr**   planations to support your answ

   (i) [**3 pts**] Consider a linear regression model with on   , $y = b$. Then given N data points $x^{(i)}$ ... $x^{(i)}$ ... re and $y^{(i)}$ is the output), minimizing the sum of squared errors resu   the median of the $y^{(i)}$ values.

   Circle one: True   False
   Brief explanation:

   (ii) [**3 pts**] Given data $\mathcal{D} = \{(\boldsymbol{x}^{(1)}, y^{(1)}), ..., (\boldsymbol{x}^{(N)}, y^{(N)})\}$, we obtain $\hat{\boldsymbol{\theta}}$, the parameters that minimize the training error cost for the linear regression model $y = \boldsymbol{\theta}^T\boldsymbol{x}$ we learn from $\mathcal{D}$.

   Consider a new dataset $\mathcal{D}_{\text{new}}$ generated by duplicating the points in D and adding 10 points that lie along $y = \hat{\boldsymbol{\theta}}^T\boldsymbol{x}$. Then the $\hat{\boldsymbol{\theta}}_{\text{new}}$ that we learn for $y = \boldsymbol{\theta}^T\boldsymbol{x}$ from $\mathcal{D}_{\text{new}}$ is equal to $\hat{\boldsymbol{\theta}}$.

   Circle one: True   False
   Brief explanation:

4. We have an input x and we want to estimate an output y using linear regression.

   Consider the dataset S plotted in Fig. 1 along with its associated regression line. For each of the altered data sets S^{new} plotted in Fig. 2, indicate which regression line (relative to the original one) in Fig. 3 corresponds to the regression line for the new data set. Write your answers in the table below.
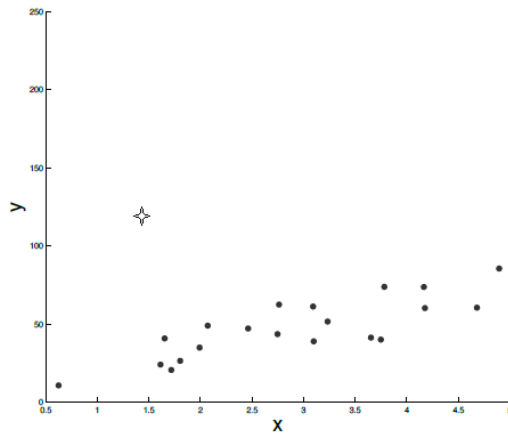
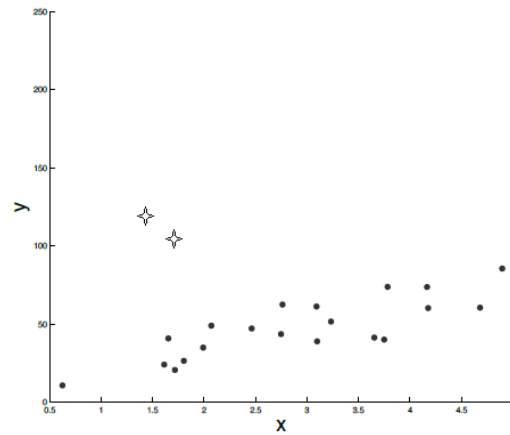 | Dataset | (a) | (b) | (c) | (d) | (e) |
   |---------|-----|-----|-----|-----|-----|
   | Regression line | | | | | |
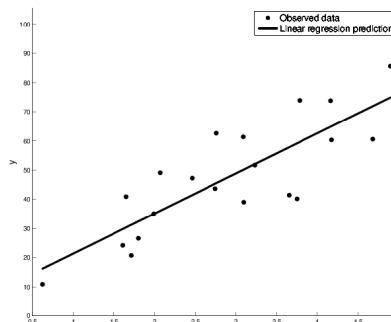
Figure 1: An observed data set and its associated regression line

(a) Adding one outlier to the
original data set.
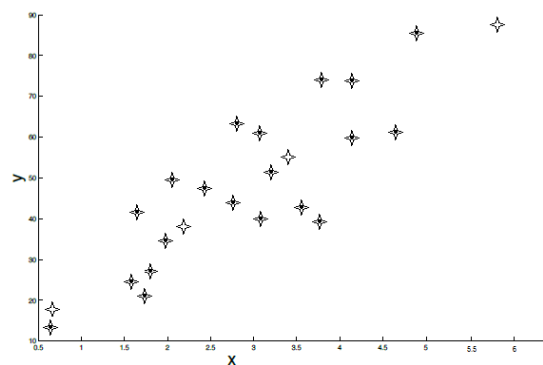

(b) Adding two outliers to the original data set.

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

(c) Adding three outliers to the original data set. Two on one side
and one on the other side.

(d) Duplicating the original data set.

(e) Duplicating the original data set and adding four points that
lie on the trajectory of the original regression line.

Figure 2: New data set S^{new}.

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Figure 3: New regression lines for altered data sets S^{new}.