CS 307
Fall 2023
DALPIAZ

Week 01
Regression

\[ Y \mid X = x \]

\[ E[Y \mid X = x] \]

mean or \( Y \) per some \( X \)

\text{Distribution of \( Y \)} \text{ for some \( X \)}

\text{Want to learn}
Data

$X \rightarrow y$

"Tabular View"

"Graphical View"
Intuition

To predict at this $X$, find other data with a similar $X$ value.
1. Fix \( k \) \( \in \) tuning parameter

2. Find the \( k \)-nearest neighbors

3. Ave \( y_j \) of neighbors

\( k \)-nearest neighbors
Linear Regression

\[ Y = \beta_0 + \beta_1 x + \epsilon \]

\[ \epsilon \sim N(0, \sigma^2) \]

**Linear Regression**

Parametric
$k = 3$

Root mean square error (RMSE) = $\sqrt{\sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2}$
$k = 1$

No errors!
A model generalized well if it is able to make good predictions for "new data."
$y = f(x) + \varepsilon$

Signal

Learn this

Noise

Get this

If you do, you have "corpit"