REUMass Amherst 2015 Data Science Bootcamp

Day 3: Classification, Regression, and Model Selection

Prof. Ben Marlin
marlin@cs.umass.edu

Plan for Day 3:

- More on Classification
- Introduction to Regression
- Model Selection and Hyperparameters

Classification

Definition: The Classification Task

Given a feature vector $\mathbf{x} \in \mathbb{R}^D$ that describes an object that belongs to one of C classes from the set $\mathcal{Y}$, predict which class the object belongs to.

Definition: Classifier Learning

Given a data set of example pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1 : N\}$ where $\mathbf{x}_i \in \mathbb{R}^D$ is a feature vector and $y_i \in \mathcal{Y}$ is a class label, learn a function $f : \mathbb{R}^D \rightarrow \mathcal{Y}$ that accurately predicts the class label y for any feature vector $\mathbf{x}$.

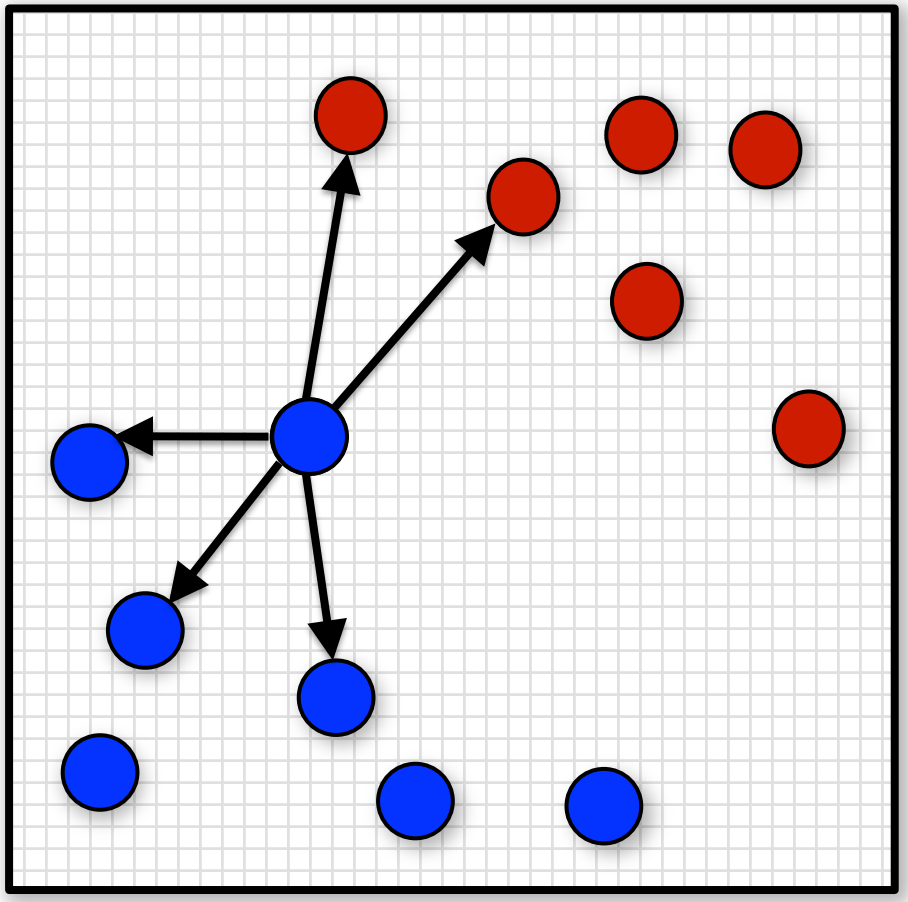
Classifier Performance

Definition: Classification Error Rate

Given a data set of example pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1 : N\}$ and a function $f : \mathbb{R}^D \rightarrow \mathcal{Y}$, the classification error rate of f on $\mathcal{D}$ is:

$$Err(f, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(y_i \neq f(\mathbf{x}_i))$$

K Nearest Neighbors



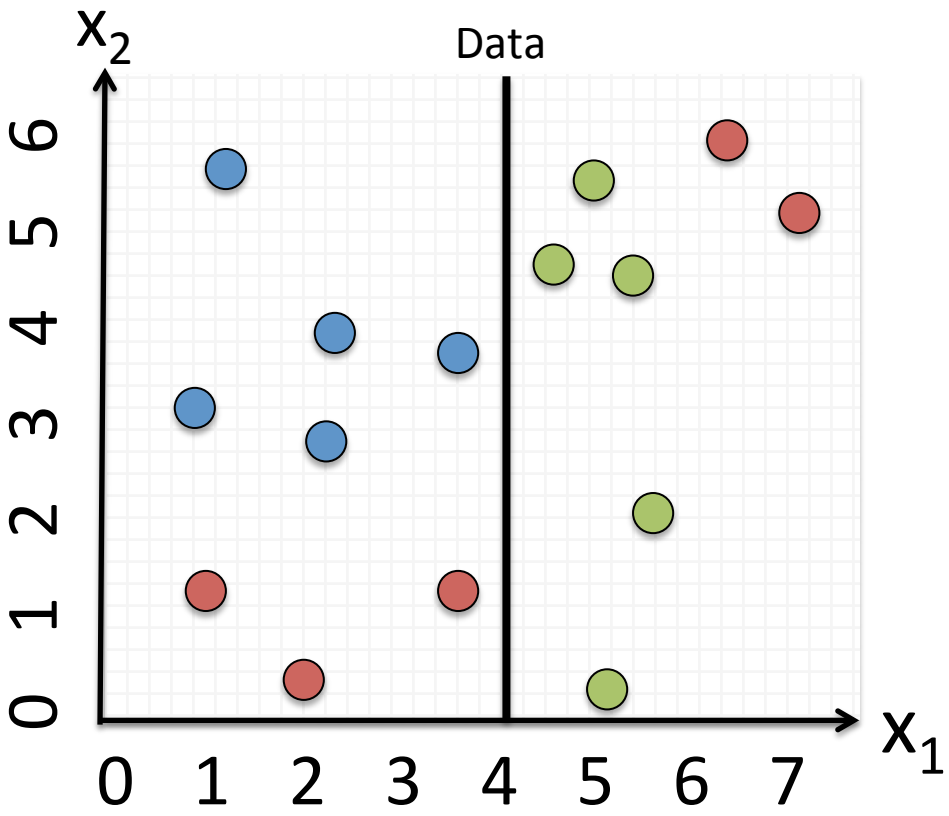
$$f_{KNN}(\mathbf{x}) = \arg\max_{y \in \mathcal{Y}} \sum_{i \in \mathcal{N}_K(\mathbf{x})} \mathbb{I}[y_i = y]$$

$$d_p(\mathbf{x}, \mathbf{x}') = ||\mathbf{x} - \mathbf{x}'||_p$$

$$= \left(\sum_{i=1}^{D} |x_d - x'_d|^p \right)^{1/p}$$

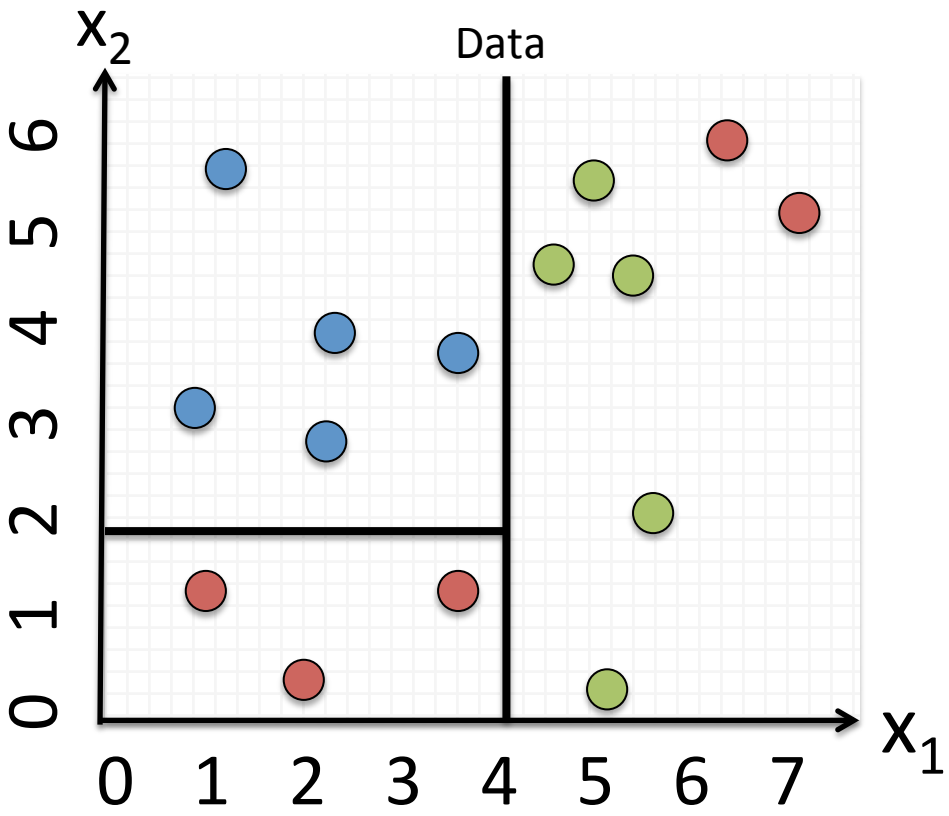
- Need to pick number of neighbors K.
- Need to define a distance function.

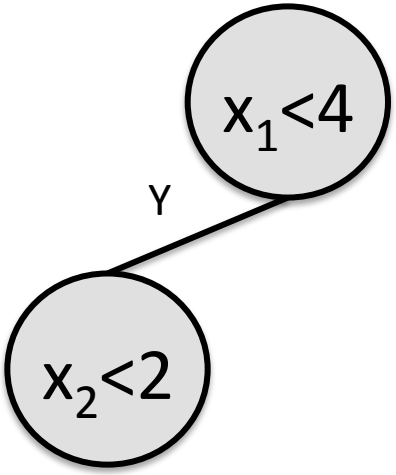
Decision Trees



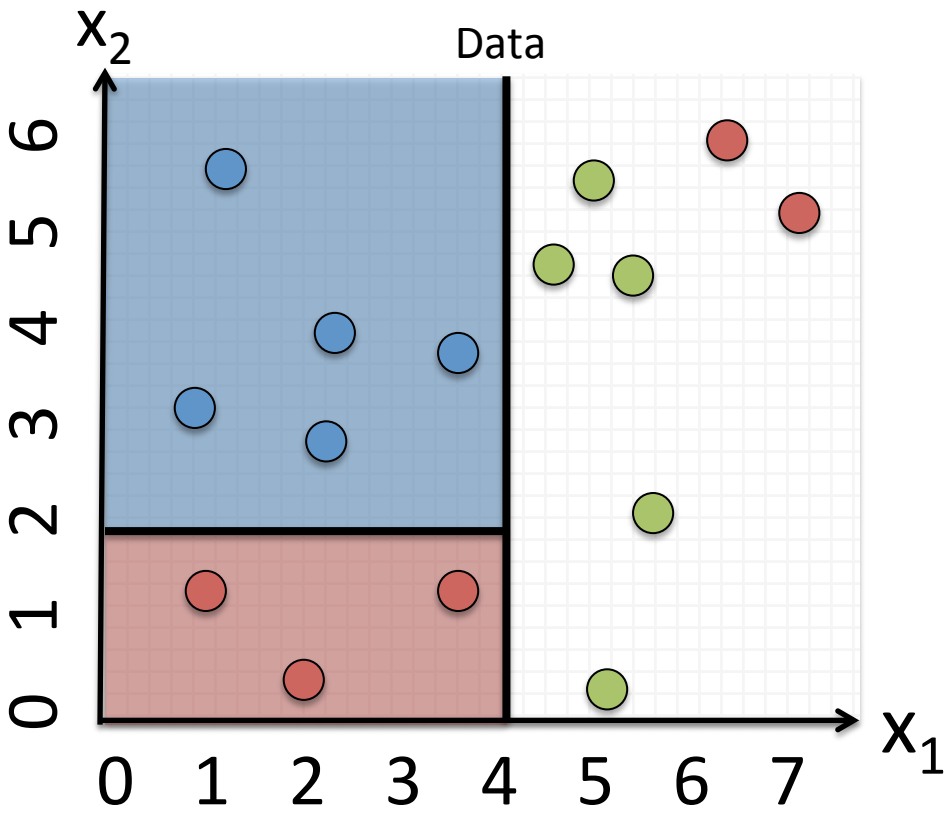
$x_1 < 4$

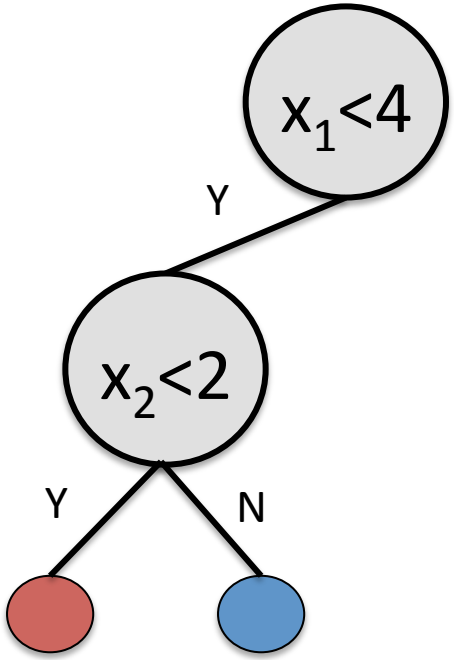
Decision Trees



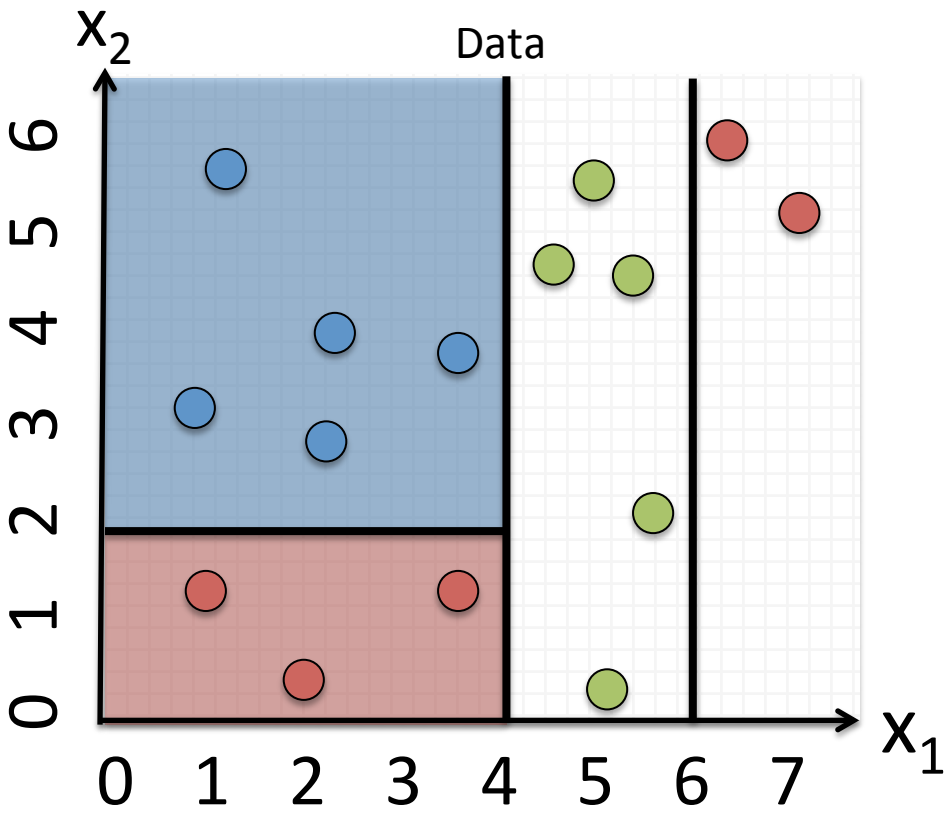


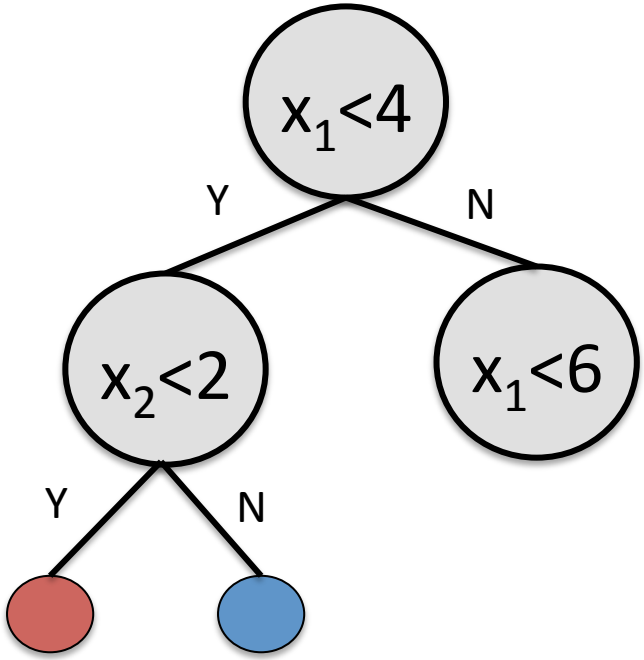
Decision Trees



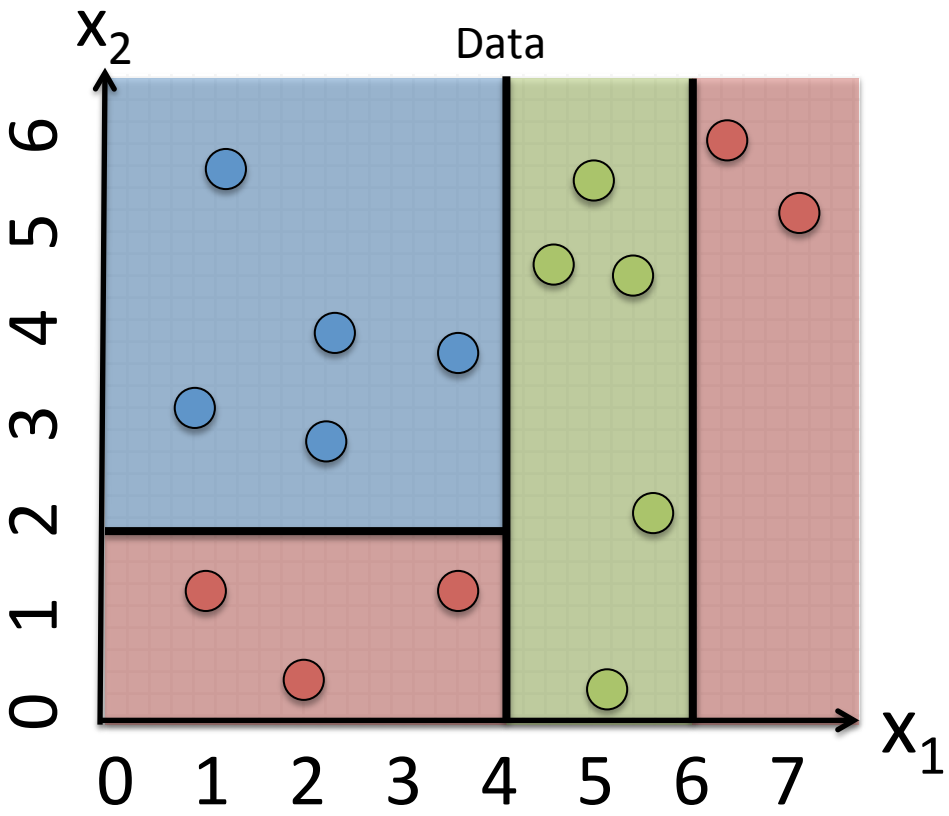


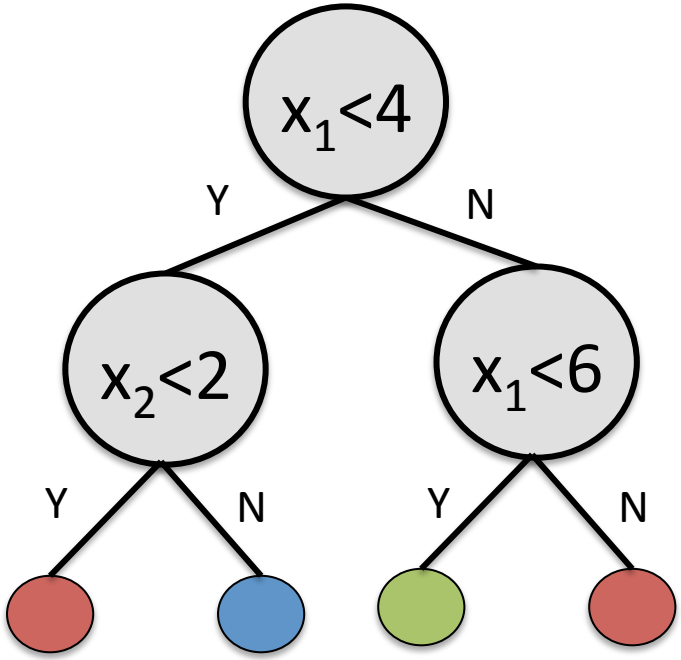
Decision Trees





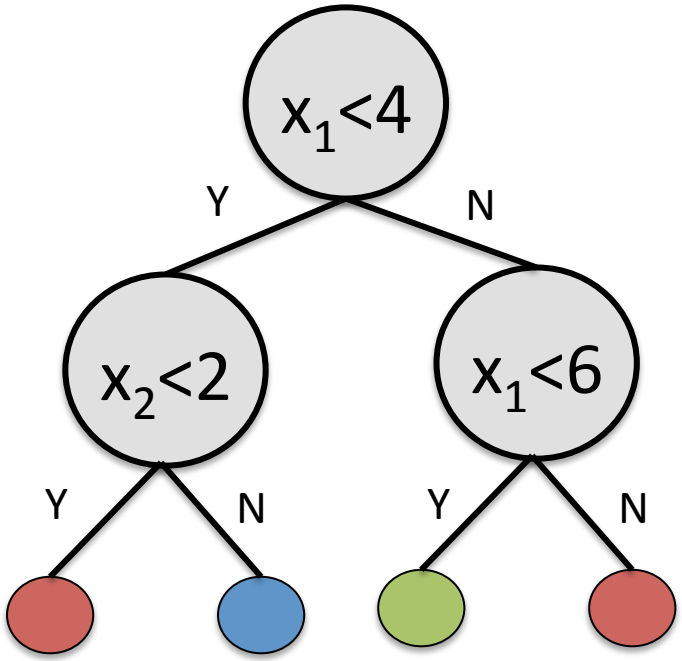
Decision Trees



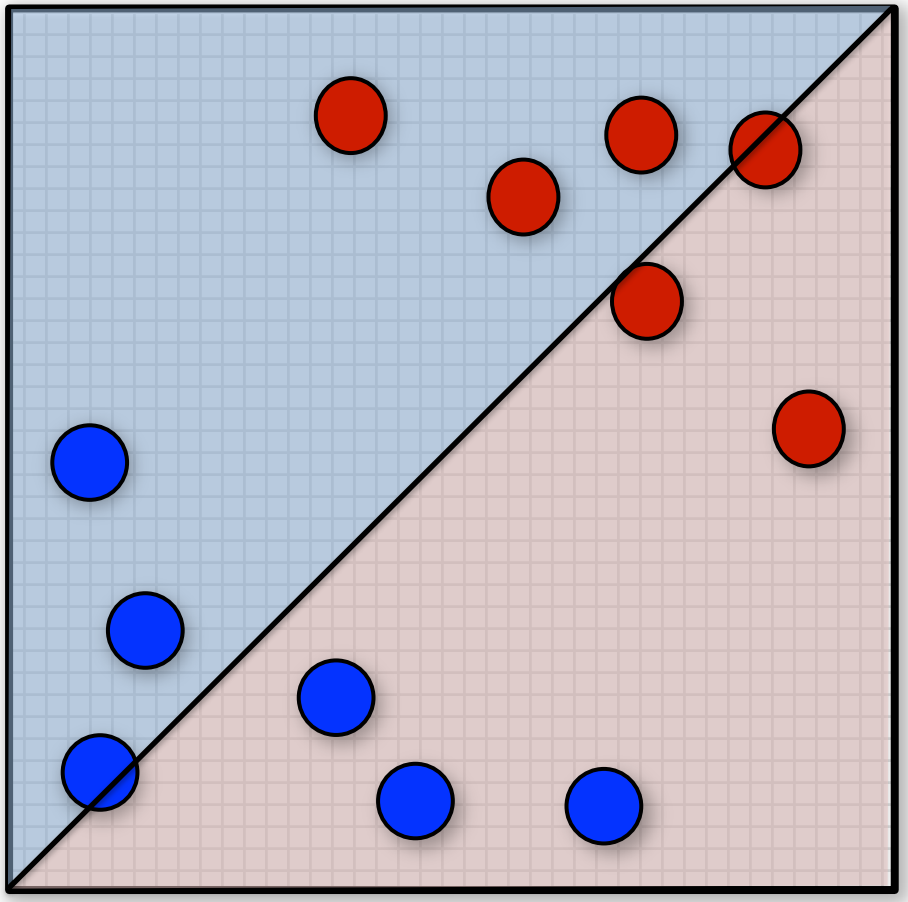


Decision Trees

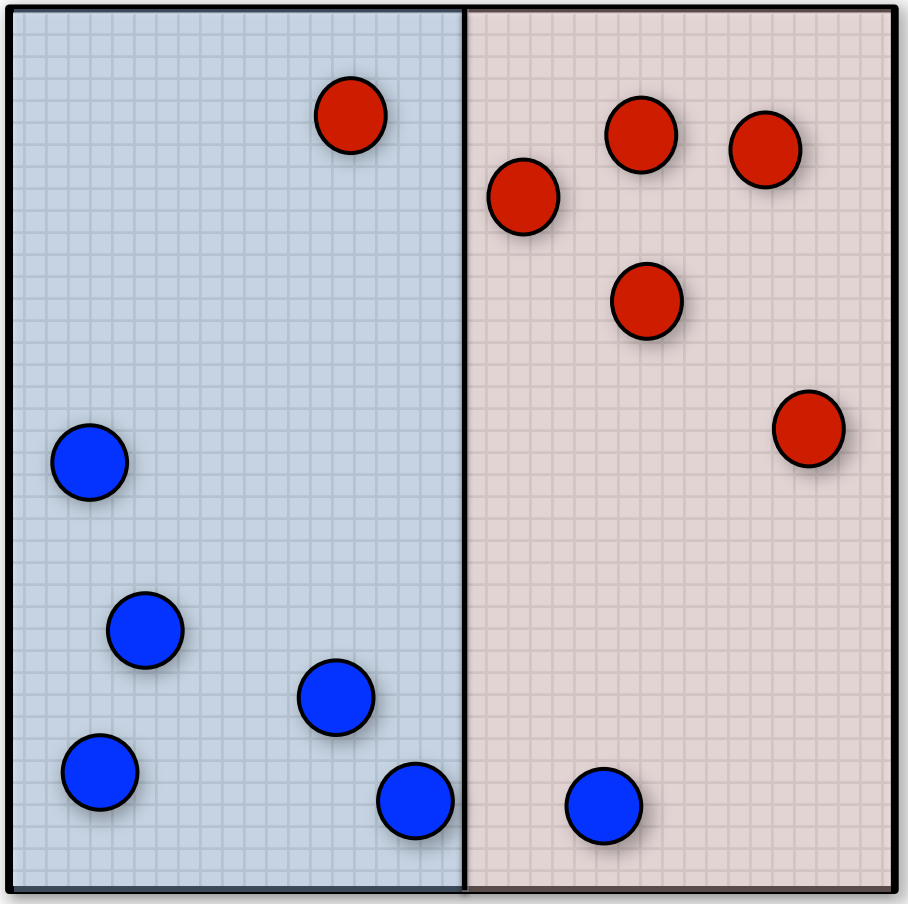
- Need to choose the maximum depth of the tree.
- Need to choose a splitting criterion.



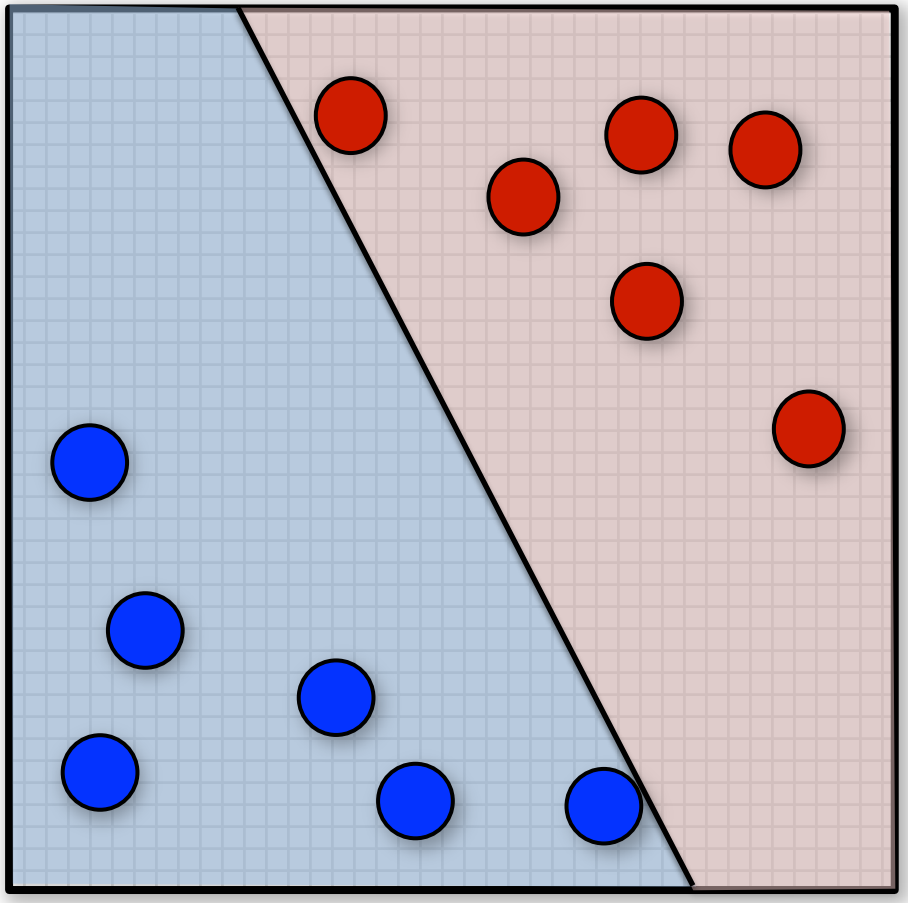
Linear Classifiers



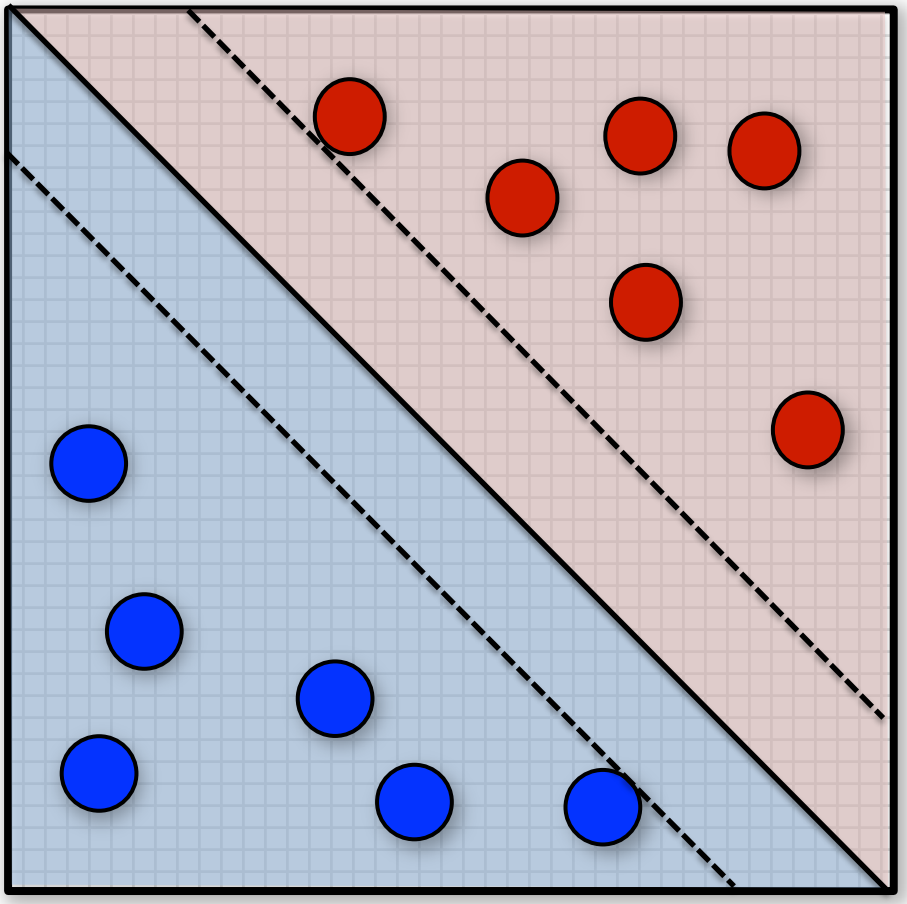
Linear Classifiers



Linear Classifiers



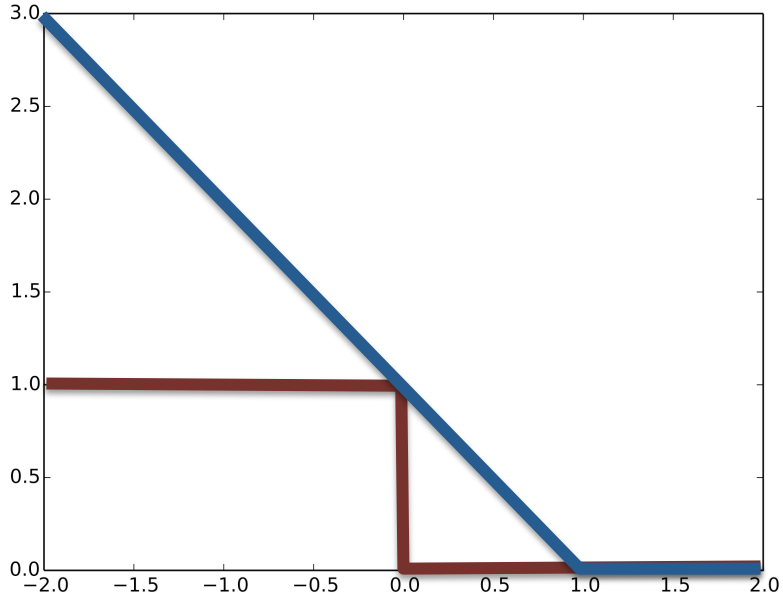
Linear Classifiers



$$f_{SVM}(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$$

$$C\sum_{i=1}^{n}\max(0, 1 - y_i \cdot g(\mathbf{x}_i)) + ||\mathbf{w}||_2^2$$

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$



Non-Linear SVMs and Basis Expansion

$$C \sum_{i=1}^{n} \max(0, 1 - y_i \cdot g(\mathbf{x}_i)) + ||\mathbf{w}||_2^2$$

$$g(\phi(\mathbf{x}_i)) = \mathbf{w}^T \phi(\mathbf{x}_j) + b$$

$$\phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \ldots, \phi_K(\mathbf{x})]$$

Non-Linear SVMs and Kernels

$$g(\phi(\mathbf{x}_i)) = \mathbf{w}^T\phi(\mathbf{x}_i) + b$$

$$= \sum_{j=1}^{N} \alpha_j \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_i) + b$$

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$$

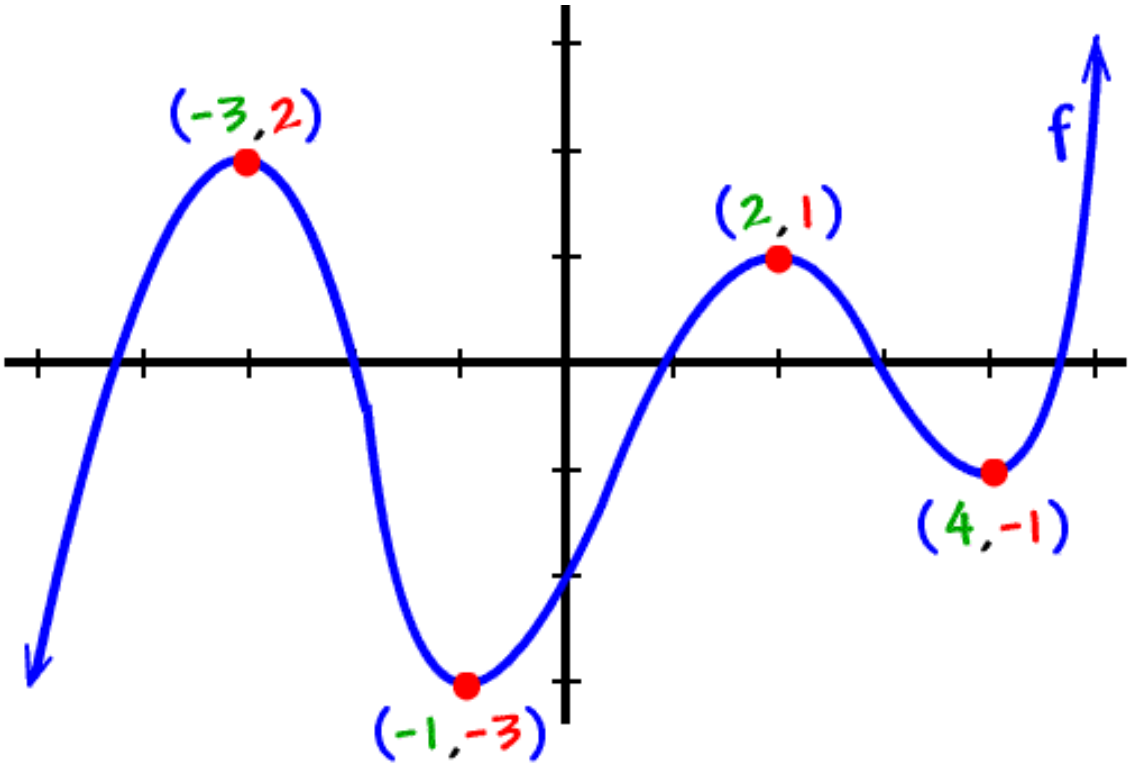
Degree B Polynomial Kernel: $\mathcal{K}_P(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T\mathbf{x}' + 1)^B$

Gaussian/RBF Kernel: $\mathcal{K}_G(\mathbf{x}, \mathbf{x}') = \exp(-\gamma||\mathbf{x} - \mathbf{x}'||_2^2)$

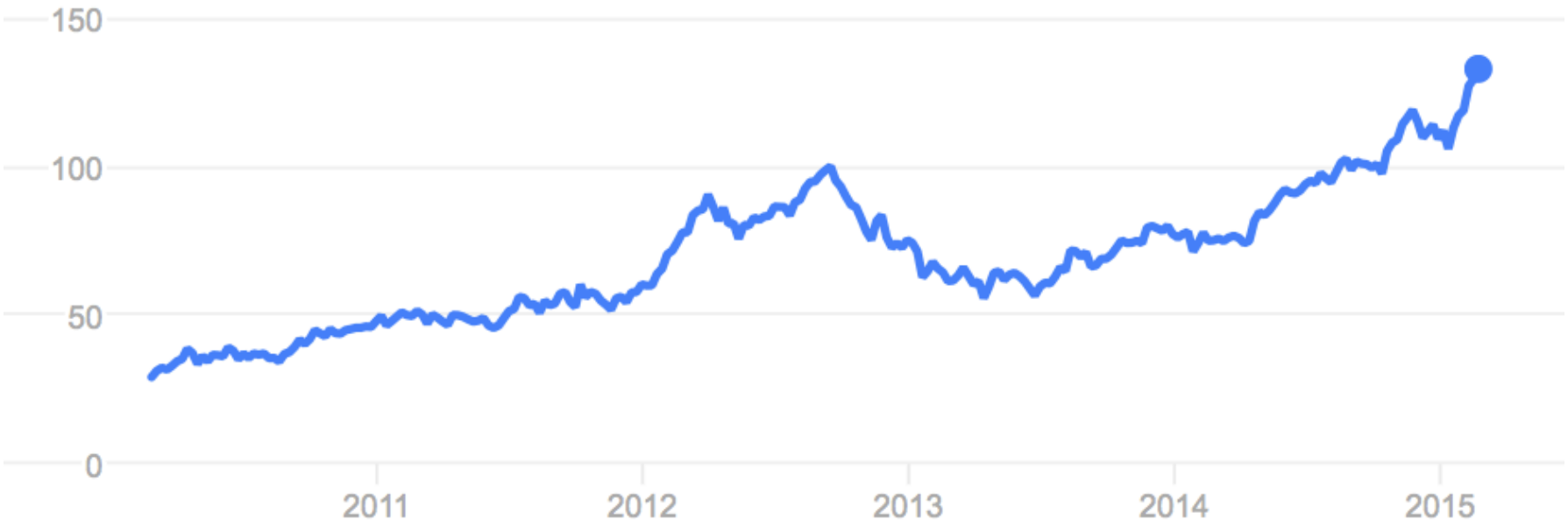
Regression

Definition: The Regression Task

Given a feature vector $\mathbf{x} \in \mathbb{R}^D$, predict it's corresponding output value y.

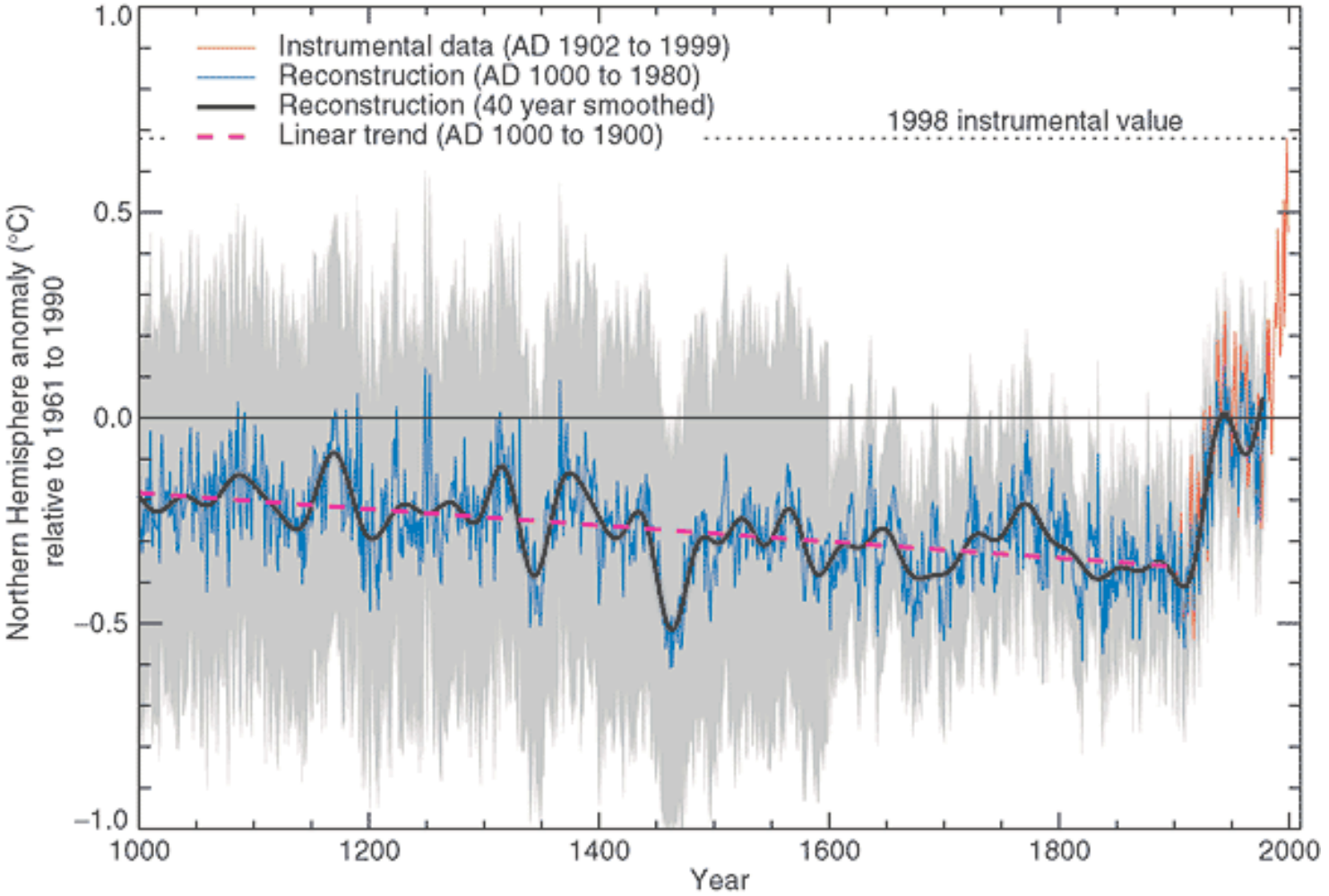


Example: Stock Prices

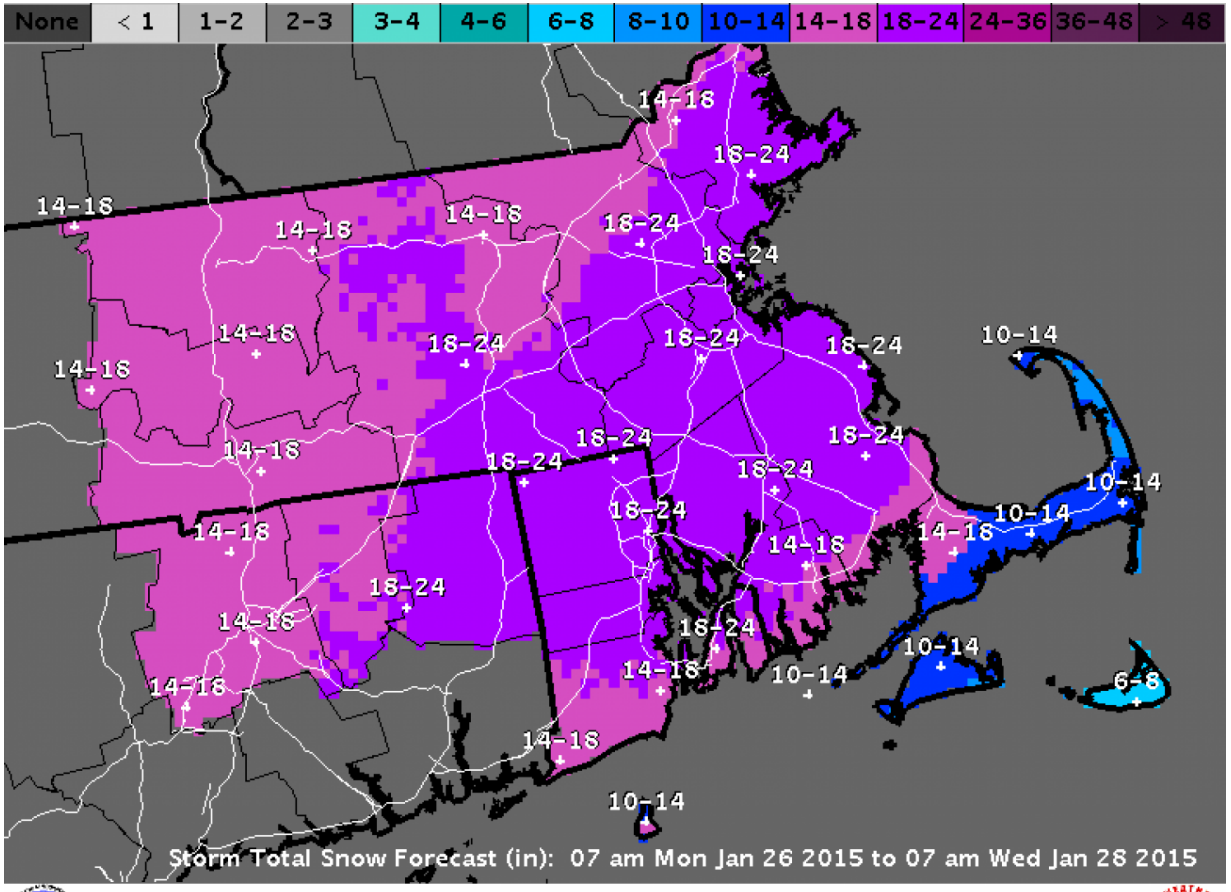


Open	130.02	Market cap	755.53B
High	133.00	P/E ratio (ttm)	17.92
Low	129.66	Dividend yield	1.41%

Example: Climate Data



Example: Weather



The Regression Learning Problem

Definition: Regression Learning Problem

Given a data set of example pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1 : N\}$ where $\mathbf{x}_i \in \mathbb{R}^D$ is a feature vector and $y_i \in \mathbb{R}$ is the output, learn a function $f : \mathbb{R}^D \rightarrow \mathbb{R}$ that accurately predicts y for any feature vector $\mathbf{x}$.

$$MAE(f, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} |y_i - f(\mathbf{x}_i)|$$

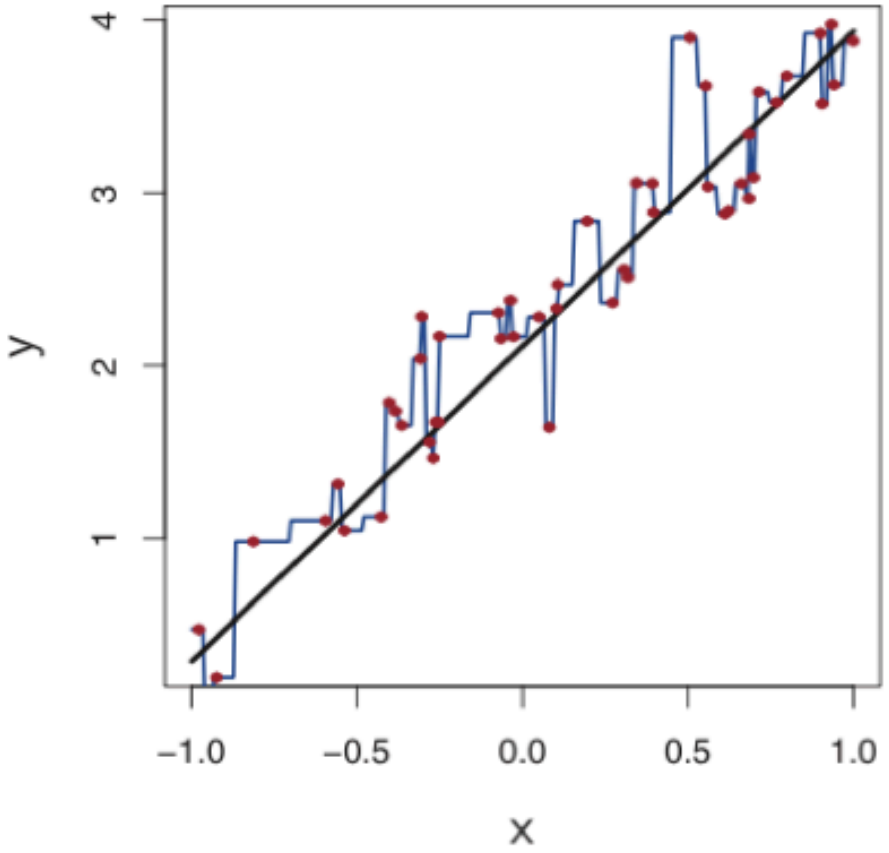
$$MSE(f, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - f(\mathbf{x}_i))^2$$

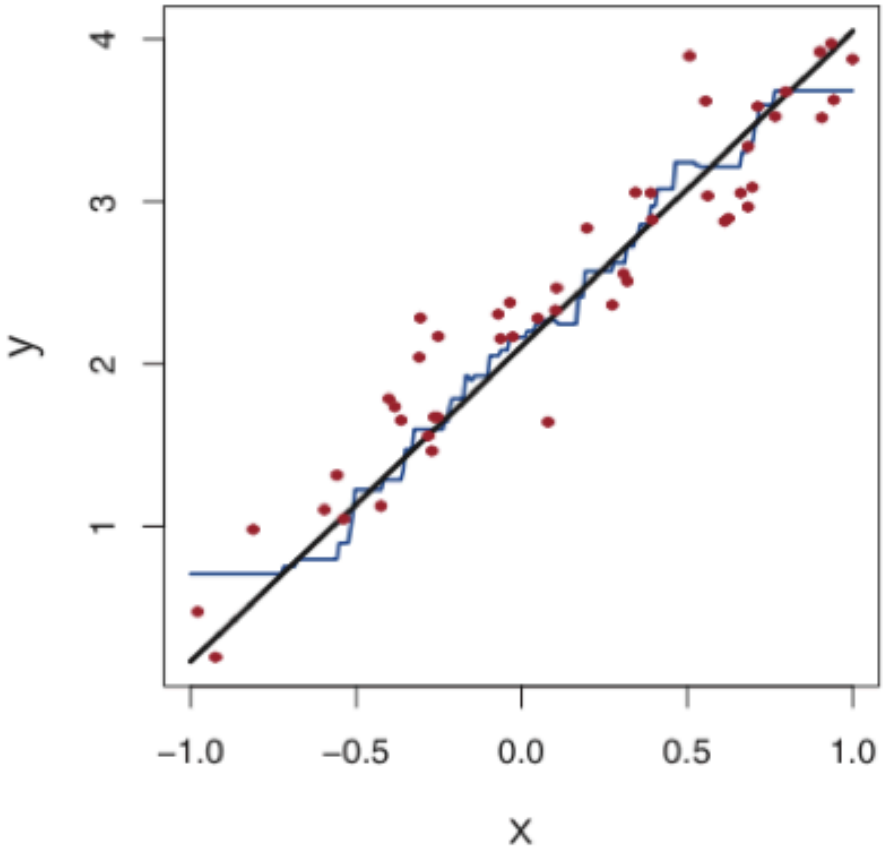
KNN Regression

KNN Regression Function

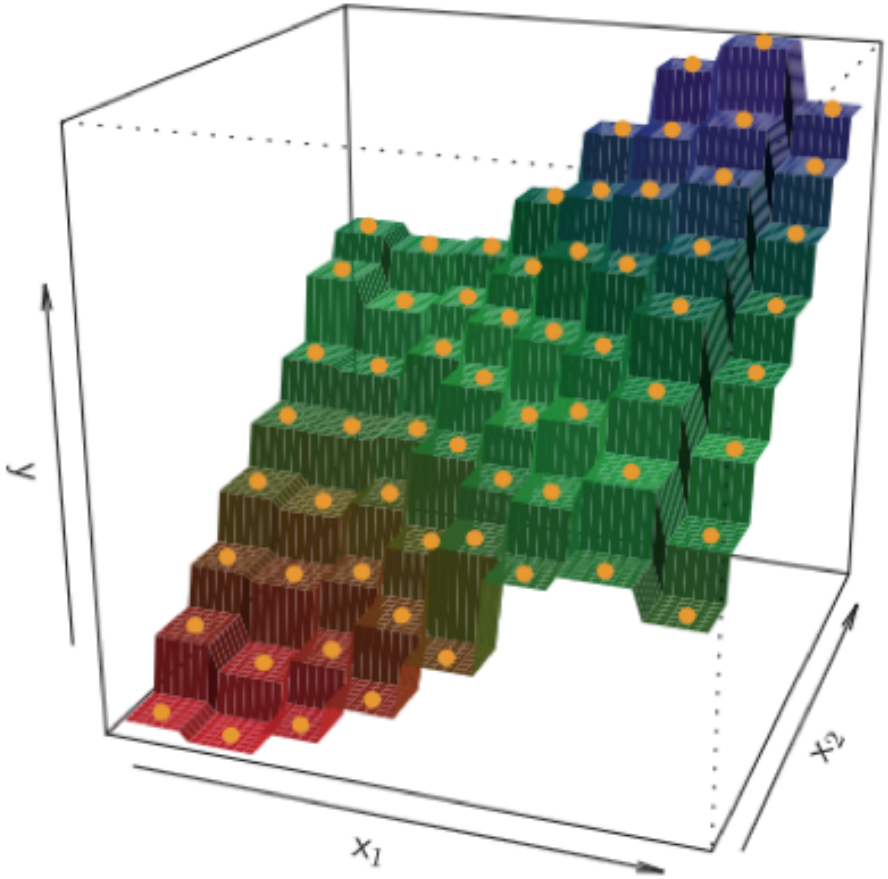
$$f_{KNN}(\mathbf{x}) = \frac{1}{K} \sum_{i \in \mathcal{N}_K(\mathbf{x})} y_i$$

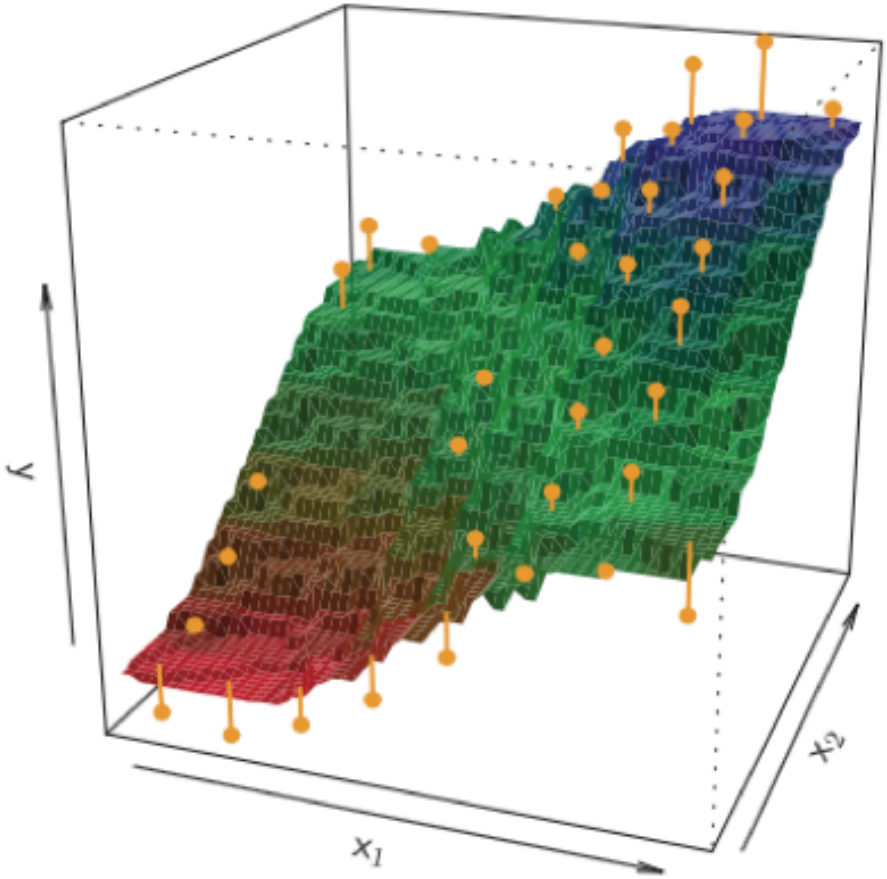
KNN Regression (K=1 vs K=9)





KNN Regression (K=1 vs K=9)





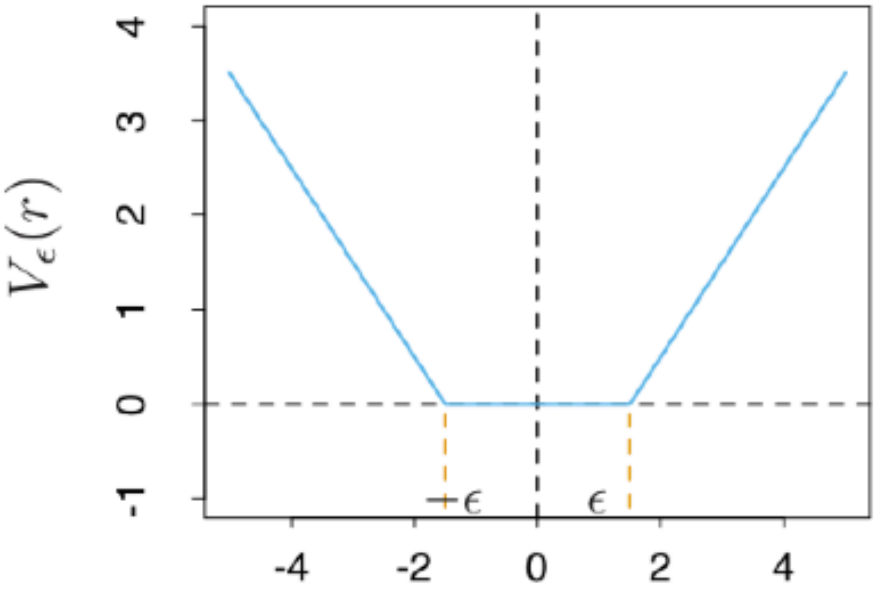
Linear Regression

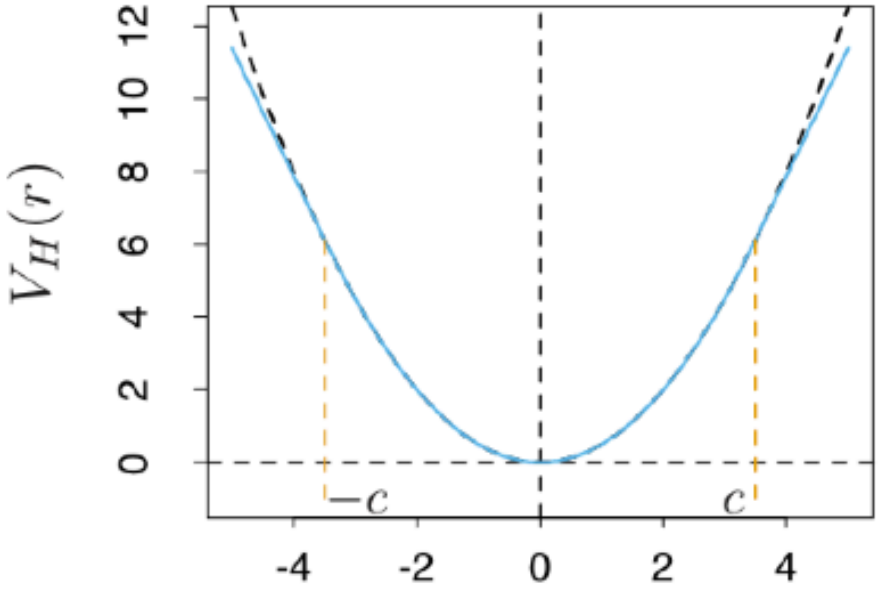
Linear Regression Function

$$f_{Lin}(\mathbf{x}) = \left(\sum_{d=1}^{D} w_d x_d \right) + b = \mathbf{x}\mathbf{w} + b$$

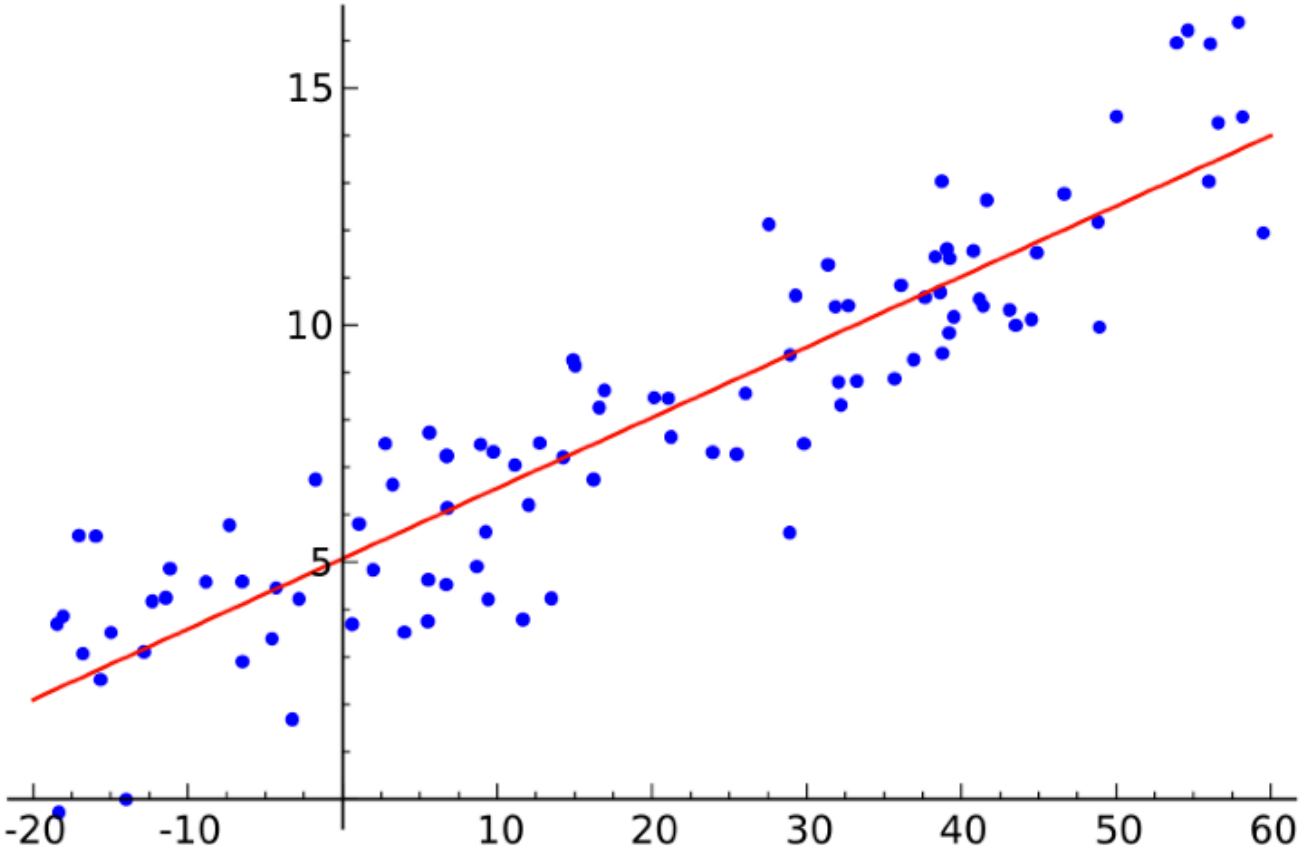
Learning Linear Regression Models

$$\mathbf{w}^*, b^* = \arg\min_{\mathbf{w},b} \frac{1}{N} \sum_{i=1}^{N} V_\epsilon(y_i - \mathbf{x}_i\mathbf{w} - b) + \lambda||\mathbf{w}||_2^2$$

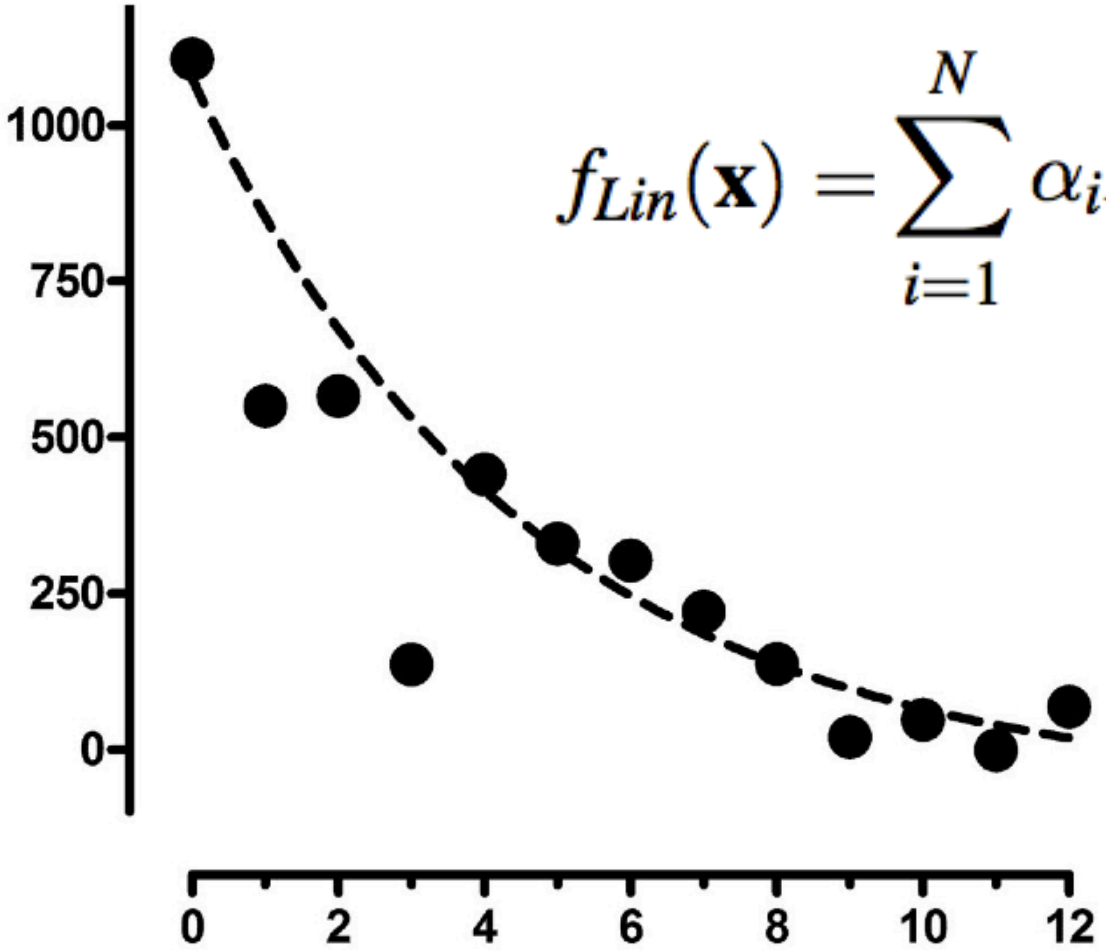




Linear Regression



Non-Linear Regression With Kernels

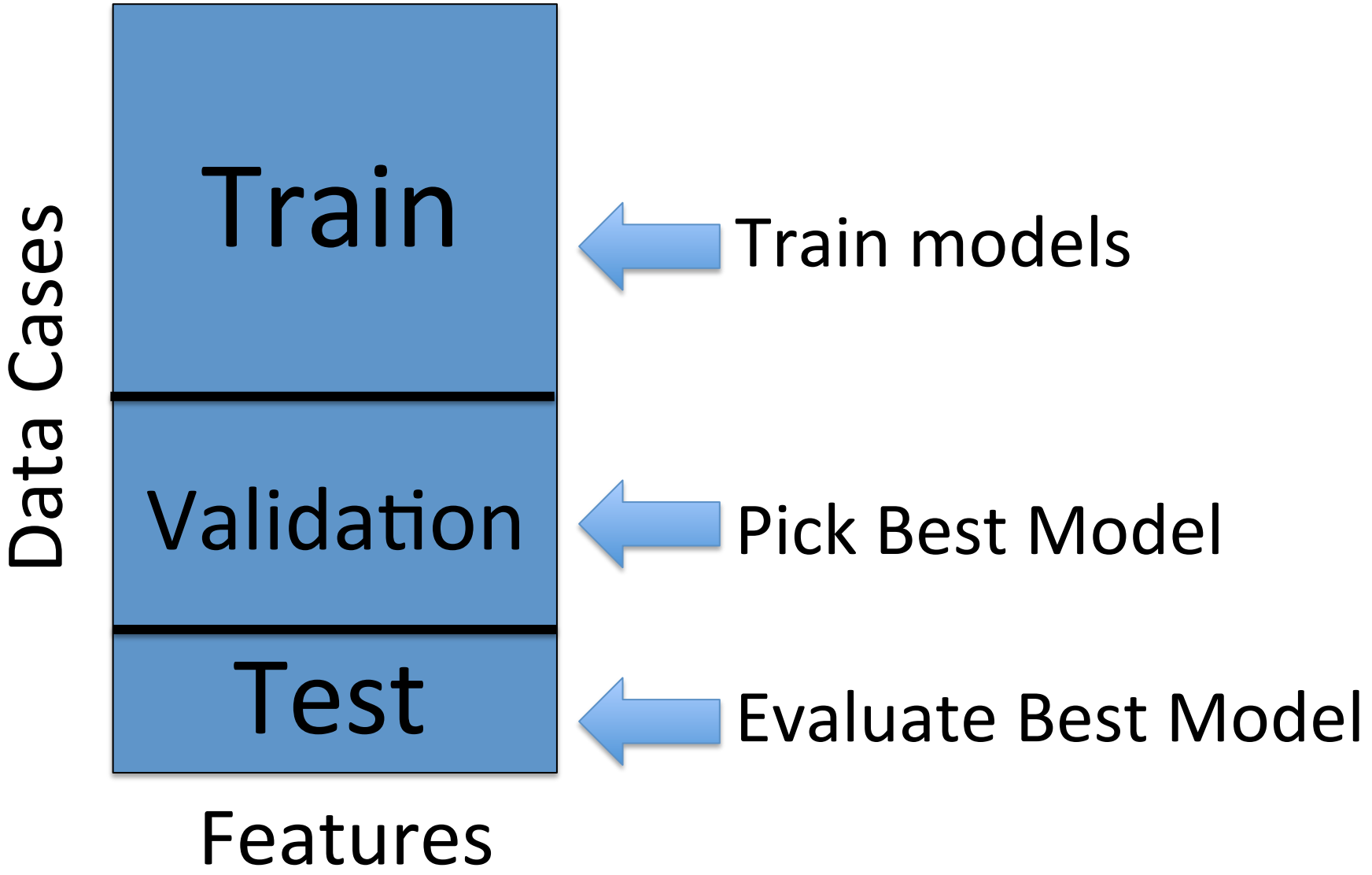


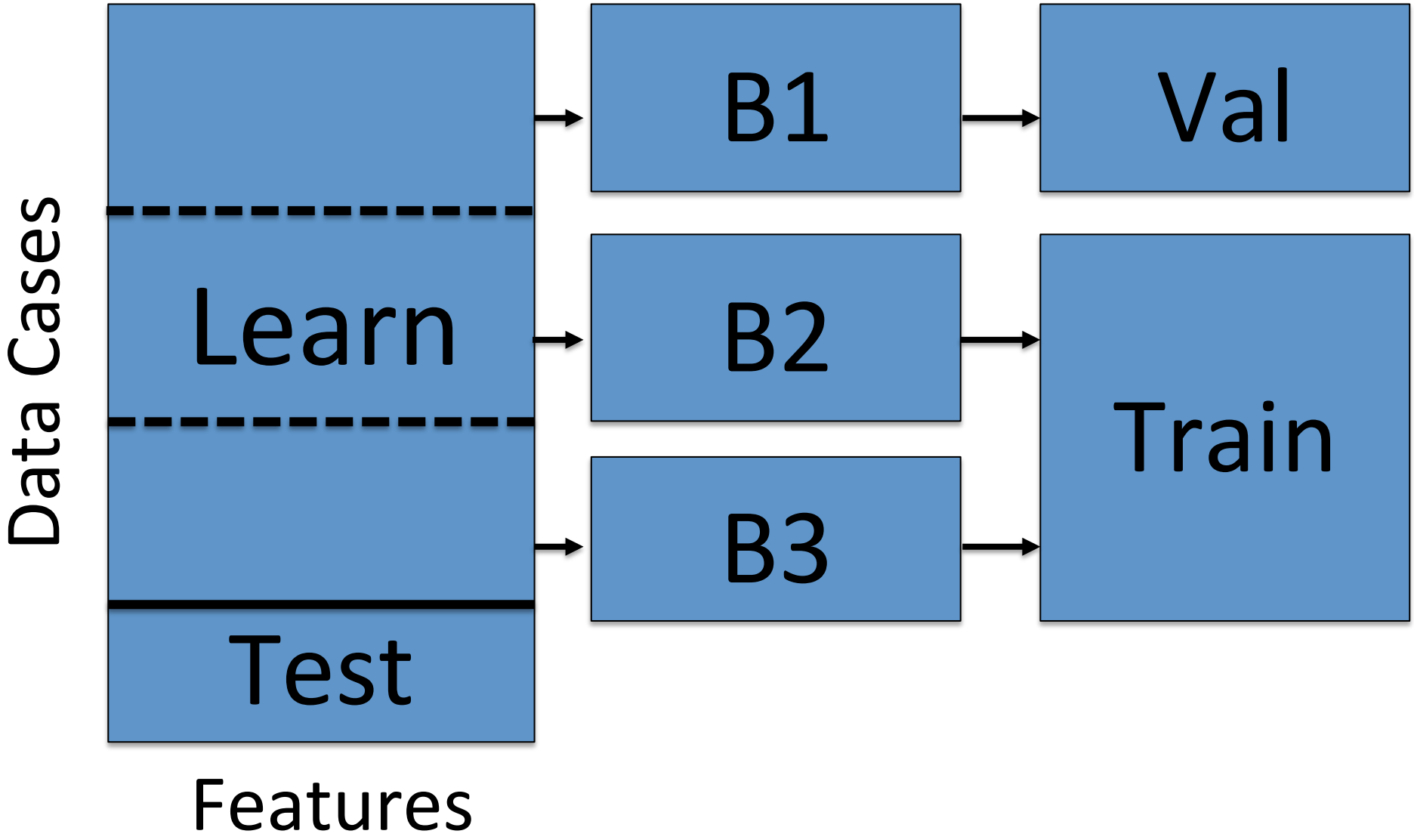
$$f_{Lin}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}, \mathbf{x}_i)$$

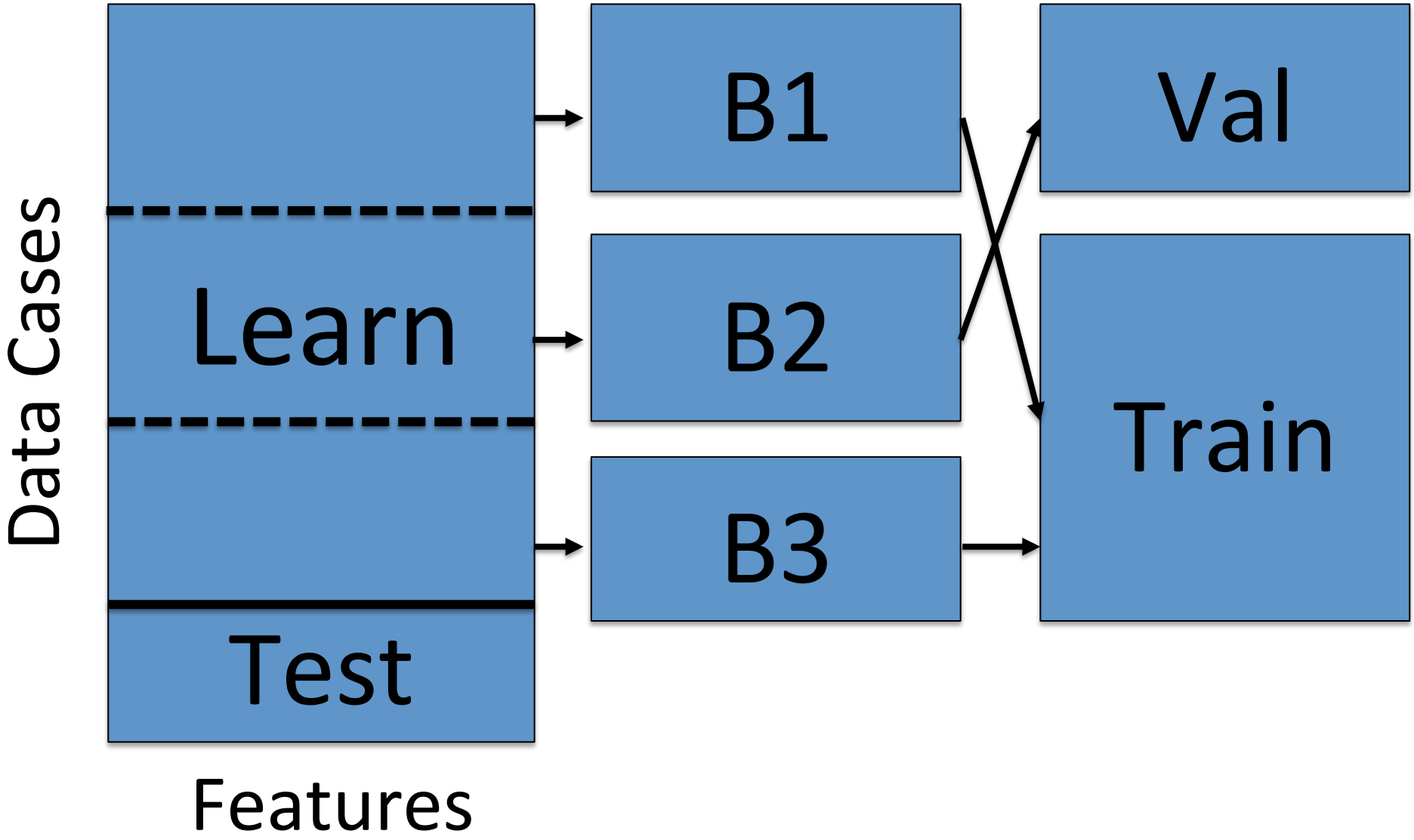
Model Selection & Hyperparameters

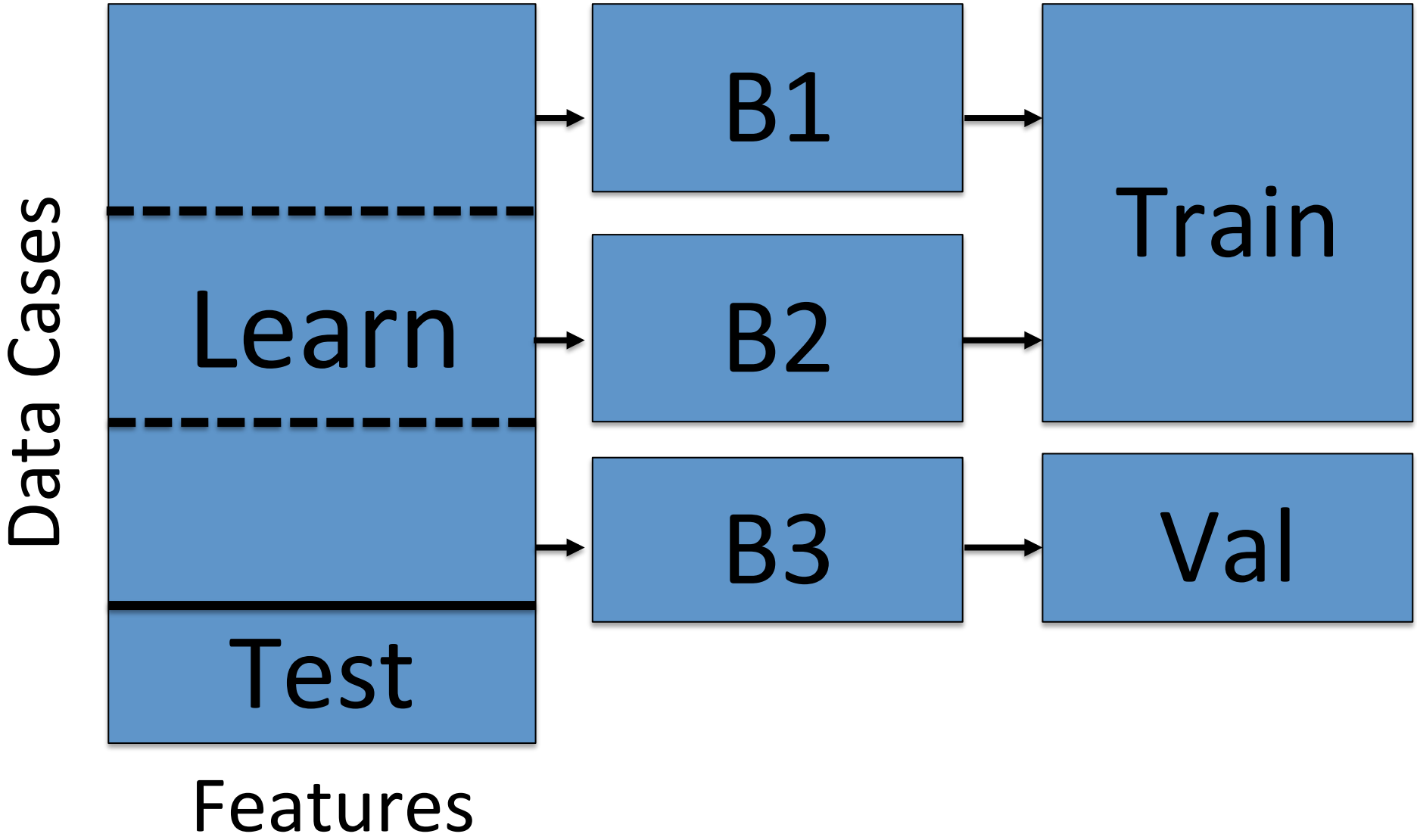
Hyperparameter and Model Selection:

- All of the models we've seen have some free parameters (called hyperparameters) that need to be chosen.
- Think of the K in KNN, the depth of a decision tree, the kernel and regularization parameter in an SVM.
- The most common methods for choosing hyperparameters in machine learning are validation set methods.


Data Cases
Train
Validation
Test
Features
Train models
Pick Best Model
Evaluate Best Model


Data Cases
Learn
Test
Features
B1
B2
B3
Val
Train


Data Cases
Learn
Test
Features
B1
B2
B3
Val
Train


Data Cases
Learn
Test
Features
B1
B2
B3
Train
Val