Syllabus on webpage : Mark Kon

3 Pillars of ML

1. Learning Theory

↳ Ex: classification & learning algorithms

2. The problem

Given an unknown function $f(x)$

learn the function from examples.

Overview of Course Statistical - Learning

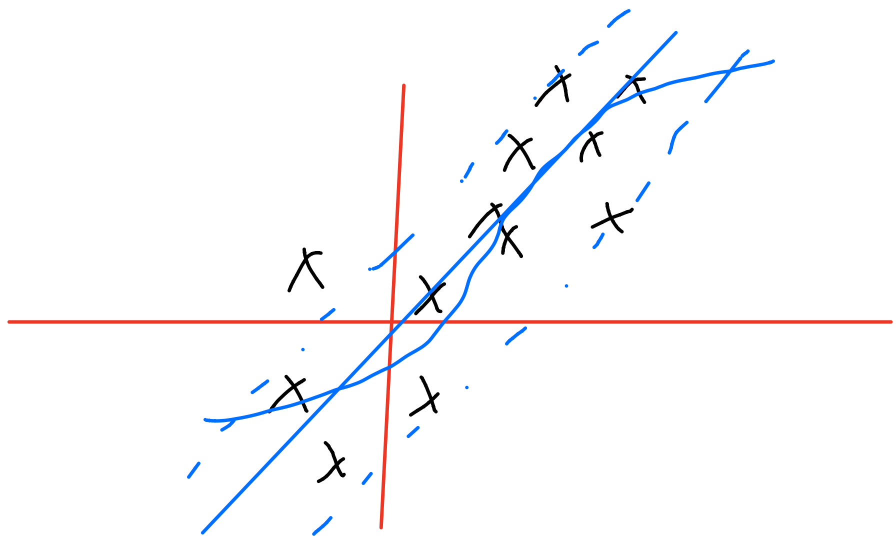
Predict Y from $\vec{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$

with

Training set $T = \{(\vec{x}_i, y_i)\}_{i=1}^{N}$

answer: $\hat{Y} = f(x)$

Ex: $f(x) = X\beta$ linear regression



$Y = f(X) + \varepsilon$

Sometimes we'll focus on the general pred. problem

$\mathbb{P}(X, Y) = \mathbb{P}(Y \mid X)\, \mathbb{P}(X)$

Our job to recover $\mathbb{P}(X, Y)$ from T.

And also $\mathbb{E}[Y|X]$

Rmk: Supervised learning: $T = (x_i, y_i)_{i=1}^n$

Unsupervised learning: $T = (x_i)_{i=1}^n$

Rmk: Y can have any structure

Ex: Nearest Neighbor Model

$Y = f(x)$ $\quad J = (x_i, y_i)_{i=1}^n$

Given X,

$$\hat{y} \overset{NN}{=} Y^* \text{ s.t. } \min_i d(x_i, x) = x^*$$

$$\overset{kNN}{=} \frac{1}{|N_{x_*}|} \sum_{N_{x_*}} Y_i$$

Ex: $\hat{y}(\vec{x}) = \sum_{i=1}^{N} w(\vec{x}-\vec{x}_i)\, y_i$

usually a Kernel.

Kernel Method

Nearest Neighbor $\subseteq$ Kernel Method $\subseteq$ Regression

$n=1$ $\quad 1 < n < N \quad$ $n=N$

Ex: Basis function methods

$$\hat{y}(x) = \sum_{m=1}^{m} \beta_m h_m(\vec{x})$$

"Spline" Models