FODA

L1 : Class Overview

Language of Data Analysis

- **Probability**

$X = \{x_1, x_2, \ldots, x_n\} \sim U$ (iid)

X: data

U: unknown distribution

$X_i = X_j$

- **Linear Algebra**

data point $x_i \in \mathbb{R}^d$

d-dimensional Euclidean space

messy data → "featurization" → $= \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,d} \end{bmatrix}$

Sum of Squared Errors

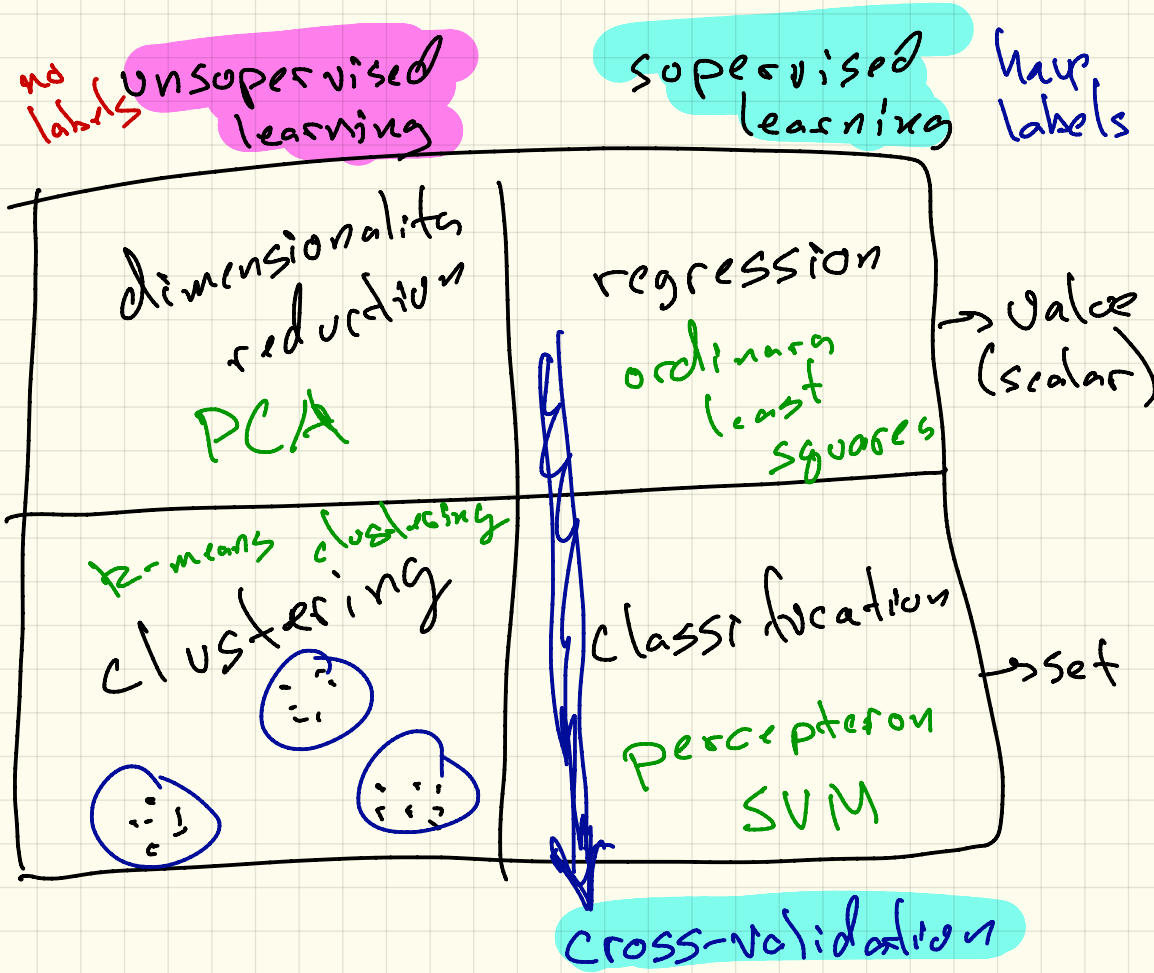
Find model M to minimize

$$\sum_{i=1}^{n} \underbrace{\left(M(x_i) - x_i\right)}_{\text{residual}}^2$$

Why?

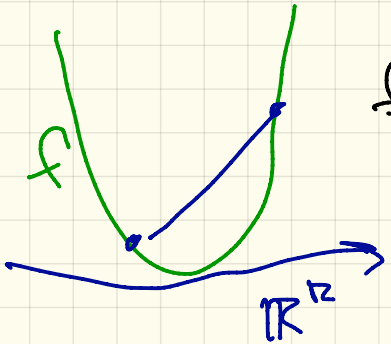
- Easy to optimize
- Bayesian Inference

Bayes' Rule + Gaussian Noise


no labels
unsupervised learning
supervised learning
have labels
dimensionality reduction
PCA
regression
ordinary least squares
value (scalar)
k-means clustering
clustering
classification
perceptron
SVM
set
cross-validation

Iterative Algorithms

$f : \{\text{space of models}\} \to \mathbb{R}^k \to \mathbb{R}$



find model M minimize $f(M)$

$M_0 \leftarrow$ initialize

$M_0 \to M_1 \to M_2 \to \dots \to M_t$

$f(M_0) > f(M_1) > f(M)_2 > \dots > f(M_t)$

Gradient Descent

$M \in \mathbb{R}^k \quad (M_\alpha, \alpha \in \mathbb{R}^k)$