Linear Regression

Aarti Singh (Instructor), HMW-Alexander (Noter)

January 30, 2017

Back to Index

Contents

Resources

- Lecture

---

1 Discrete to Continuous Labels

From classification to regression

1.1 Task

Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$, Construct prediction rule $f : \mathcal{X} \to \mathcal{Y}$

1.2 Performance Measure

- Quantifies knowledge gained.
- Measure of closeness between true label Y and prediction f(X)
  - 0/1 lose:$loss(Y, f(X)) = 1_{f(X) \neq Y}$. Risk: probability of error
  - square loss: $loss(Y, f(X)) = (f(X) - Y)^2$. Risk: mean square error
- How well does the predictor perform on average?

$$Risk\ R(f) = \mathbb{E}[loss(Y, f(X))],\ (X, Y) \sim P_{XY}$$

1.3 Bayes Optimal Rule

- ideal goal: Construct prediction rule $f^* : \mathcal{X} \to \mathcal{Y}$

$$f^* = \arg\min_f E_{XY}[loss(Y, f(X))]$$

  (Bayes optimal rule)

- Best possible performance:

$$\forall f,\ R(f^*) \leq R(f)$$

  (Bayes Risk)

Problem: P_{XY} is unknown.
Solution: Training data provides a glimpse of P_{XY}

$$\text{(observed) } \{(X_i, Y_i)\} \sim_{i.i.d} P_{XY} \text{ unknown}$$

2 Macine Learning Algortihm

- Model based approach: use data to learn a model for P_{XY}
- Model-free approach: use data to learn mapping directly

2.1 Empirical Risk Minimization (model-free)

- Optimal predictor:

$$f^* = \arg\min_f \mathbb{E}[(f(X) - Y)^2]$$

- Empirical Minimizer:

$$\hat{f}_n = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (f(X) - Y)^2$$

$\mathcal{F}$ is the class of predictors:

- Linear
- Polynomial
- Nonlinear

3 Linear Regression

$$f(\vec{X}) = \sum_{i=0}^{p} \beta_0 X^i = \vec{X}^T \vec{\beta},\ where\ X^0 = 1,\ \vec{\beta} = [\beta_0, \ldots, \beta_p]$$

$$\hat{\vec{\beta}} = \arg\min_{\vec{\beta}} (A\vec{\beta} - Y)^T (A\vec{\beta} - Y)$$