Machine Learning Topic 1

Introduction and Binary Classification

Aimee Barciauskas

29 March 2016

Binary Classification

- Definitions μ and η: *slide 4*
- Definition of $R(g)$ as function of $g(X)$ and of expected loss: *slide 5*
- Definition of **Bayes Classifier** $g^*(x)$: *slide 5*
  - Bayes Classifier Theorem: *bottom of slide 5*
  - Proof Bayes classifier is optimal: *slide 6*
- Definition of **plug-in classifier**: *slide 7*
- Data-based classifier and it's risk: *slide 8*
- Consistency, universal and strong: *slide 9*

Notes

- How do we classify? Pick whichever class-conditional density is bigger!
- When you expand the variance of $R(g) = \mathbb{E}[R(g) - R_n(g)]^2$, expanding the expected value reveals only terms in $R(g)$ and you find that the variance is small when $R(g)$ is small

Definitions

$\eta(x) = \frac{\mathbb{P}\{Y=1\}f_1(x)}{f(x)}$, where $f_1(x)$ is the conditional distribution of x given $Y = 1$

$f(x|Y = 1) = \frac{f(x)\mathbb{P}(Y=1|X=x)}{\mathbb{P}(Y=1)}$

$\mathbb{P}\{Y = 1\} = \mathbb{E}[\eta(x)]$

$\mathbb{P}\{X \in A\} = \mathbb{P}\{Y = 1\}f_1(x|Y = 1) + \mathbb{P}\{Y = 0\}f_0(X|Y = 0)$

Risk

$R(g) = 1 - \mathbb{E}[\mathbb{I}_{g(X)=1}\eta(X)] - \mathbb{E}[\mathbb{I}_{g(X)=0}(1 - \eta(X))]$

$R(g_n) = \frac{1}{n}Bin(n, R(g))$

$\mathbb{E}[R(g_n)] = \frac{1}{n}(nR(g)) = R_g$

$\bar{R}$: best risk of family

$\hat{R_n}$: empirical risk

Bayes Risk

$R^* = \inf_{g:\mathcal{R}^d \to \{0,1\}} \mathbb{P}\{g(X) \neq Y\}$

$= \mathbb{E}[min(\eta(x), 1 - \eta(x))]$

$= \frac{1}{2} - \frac{1}{2}\mathbb{E}\Bigg\{ |2\eta(X) - 1| \Bigg\}$

$= \int min(\eta(x), 1 - \eta(x))f(x)dx$ if X has density $f(x)$

$= \int min((1-p)f_0(x), pf_1(x))dx$ if X has class-conditional densities $f_i(x)$

$R^* = 0$ when $\eta \in \{0, 1\}$ everywhere

Error Estimation

Error-counting estimator:

(8.1) Given, m is the **testing** sequence:

$$\hat{L}_{n,m} = \frac{1}{m}\sum_{j=1}^{m}\mathbb{I}_{g_n(X_{n+j}) \neq Y_{n+j}}$$

Clearly unbiased: $\mathbb{E}\big\{\hat{L}_{n,m}|D_n\big\} = L_n$

The conditional distribution of $\hat{L}_{n,m}$ given D_n is binomial with paramters m and L_n