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 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*
- $\bar{R}$: best risk of family, $\hat{R_n}$: empirical risk, both being used to bound $R(g_n)$

Additional and Alternative Definitions

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

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

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

Error Estimation

Error-counting estimator:

(8.1 of Book)

Given, m is the testing sequence:

$$\hat{L_{n,m}} = \frac{1}{m} \sum_{j=1} m \not\Vdash_{g_n(X_{n+j}) \neq Y_{n+j}}$$

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

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