CO902
Supporting Lab Notes:
Gaussian Class Conditionals for Classification
Lab 4 — 29 Jan 2012

1 Optimal Classifier Shortcut

For K-class classifier problem based on $\{\mathbf{X}_i, Y_i\}$, $i = 1, \ldots, N$, D-dimensional $\mathbf{X}_i \in \mathcal{X}$, scalar $Y_i \in \{1, 2, \ldots, K\}$, the class notes showed that the optimal decision rule reduces to implimenting this RHS

$$\operatorname*{argmax}_k P(Y = k|\mathbf{X} = \mathbf{x}) \quad = \quad \operatorname*{argmax}_k p(\mathbf{x}|Y = k)P(Y = k)$$

For Gaussian data, i.e. $\mathbf{X}|Y = k \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, the class conditional distributions are

$$p(\mathbf{x}|Y = k) \quad = \quad \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right)$$

If we are using "Naive Bayes" however, then we asssume independence, and $\boldsymbol{\Sigma}_k = \text{diag}(\sigma^2_{k1}, \ldots, \sigma^2_{kD})$. "Training" then, simply consists of finding estimates $\hat{\mu}_{kd}$ and $\hat{\sigma}^2_{kd}$ for each variable $d = 1, \ldots, D$ in each class $k = 1, \ldots, K$.

Prediction is done with the (independence) conditional class distribution with estimated parameters. The log conditional class distribution, dropping constants, for a new observations is then

$$\log p(\mathbf{x}^{\text{new}}|Y = k) \quad \propto \quad -\frac{1}{2}\sum_d \log(\hat{\sigma}^2_{kd}) - \frac{1}{2}\sum_d \frac{(x^{\text{new}}_d - \hat{\mu}_{kd})^2}{\hat{\sigma}^2_{kd}}$$

What do you do if $\hat{\sigma}_{kd} = 0$ for some k, d?

TEN / January 29, 2013