EC500 - Introduction to Learning From Data

Problem Set 2

Mikhail Andreev

October 19, 2015

1 Soft Thresholding

Given:
$X = Y + Z$
$Y \perp Z$
$Z \sim N(0, \sigma^2), \quad \sigma > 0$
$Y \sim \pi(y) = 0.5\lambda \exp(-\lambda|y|), \quad \lambda > 0$

Solution:
$h_{MAP}(x) = \arg\max_y P_{y|x}(y|x)$

$P_{y|x}(y|x) = \frac{P_{y|x}(x|y)P(y)}{P(x)}$

p(x) is not related to y, so it can be ignored in the maximization.

$h_{MAP}(x) = \arg\max p(x|y)p(y)$

$p(x|y) = y + z$ for any y

Since z = zero-mean Gaussian variable, y+z is a Gaussian variable with mean y, and pdf centered at y.

$p(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-0^2}{2\sigma^2})$

$p(x|y) = \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{(x-y)^2}{2\sigma^2})$

$h_{MAP}(x) = \arg\max \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{(x-y)^2}{2\sigma^2})p(y)$

Taking the log we get:

$h_{MAP}(x) = \arg\max[\log(\frac{1}{\sigma\sqrt{2\pi}}) + \log(\exp(\frac{(x-y)^2}{2\sigma^2})) + \log(p(y))]$

Since $\log(\frac{1}{\sigma\sqrt{2\pi}})$ is a constant, we can ignore it and simplify to:

$h_{MAP}(x) = \arg\max[\frac{(x-y)^2}{2\sigma^2} + \log(p(y))]$

$p(y) = 0.5\lambda \exp(-\lambda|y|)$

$\log(p(y)) = \log(0.5\lambda) + \log(\exp(\lambda|y)) = \log(0.5\lambda) + \lambda|y|$

Since $\log(0.5\lambda)$ is again a constant across all y, we can ignore it to get:

$h_{MAP}(x) = \arg\max[\frac{(x-y)^2}{2\sigma^2} + \lambda|y|]$

To solve the equation, we take the derivative with respect to y, and set it to 0:

$\frac{y-x}{\sigma^2} + \lambda sign(y) = 0$

$y - x + \lambda\sigma^2 sign(y) = 0$

$x - \lambda\sigma^2 sign(y) = y$

$$y(x) = \begin{cases} x + \lambda\sigma^2 & x < -\lambda\sigma^2 \\ 0 & -\lambda\sigma^2 \leq x \leq \lambda\sigma^2 \\ y - \lambda\sigma^2 & x < \lambda\sigma^2 \end{cases}$$

2 QDA vs LDA

Given: Binary Classification Problem
Class Labels: y = 0, y = 1, equally likely
Feature vector: $x = (x_1, x_2)^T \in R^2$

$$p(x|y=0) = \begin{cases} \frac{1}{Z_0} & if||x||_1 \leq \frac{1}{\sqrt{2}} \\ 0 & otherwise \end{cases}$$

$$p(x|y=1) = \begin{cases} \frac{1}{Z_1} & if||x-(\sqrt{2},0)^T||_1 \leq \sqrt{2} \\ 0 & otherwise \end{cases}$$

2.1 Part a

Since y = 0 and y = 1 are equally likely, we need to find the place where the areas of x intersect for both labels. This region has an area of: $\frac{\sqrt{2}}{2}\frac{\sqrt{2}}{2} = \frac{1}{2}$.

The total area of the $p(x|y=0) \neq 0$ is 2, and the total area of the $p(x|y=0) \neq 0$ is 4.

This gives us:

$\frac{1}{Z_0} = \frac{\frac{1}{2}}{2} = \frac{1}{4}$

$\frac{1}{Z_1} = \frac{\frac{1}{2}}{4} = \frac{1}{8}$

So: $Z_0 = 4$, $Z_1 = 8$

2.2 Part b

$h_{MPE} = \arg\max p(y|x)$

$h_{MPE} = \arg\min[-\log(p(x|y)) - \log(p(y))]$

$h_{MPE} = \min(-\log(\frac{1}{Z_0} - \log\frac{1}{2}), -\log(\frac{1}{Z_1} - \log\frac{1}{2})$

$-\log(\frac{1}{Z_0} - \frac{1}{2}) = 2.079$

$-\log(\frac{1}{Z_1} - \frac{1}{2}) = 2.773$

So, $h_{MPE} = 2.079$, classified as class 0.

2.3 Part c

$P_{RISK} = P(y=0)\int_{R_2} p(x|y=0)dx + P(y=1)\int_{R_1} P(x|y=1)dx$

$P_{RISK} = \frac{1}{2}\int_{R_2} \frac{1}{Z_0}dx + \frac{1}{2}\int_{R_1} \frac{1}{Z_1}dx$

$P_{RISK} = \frac{1}{2Z_0}x|_{R_2} + \frac{1}{2Z_1}x|_{R_1}$

$P_{RISK} = \frac{1}{8Z_0} + \frac{1}{8Z_1}$

$P_{RISK} = \frac{Z_0+Z_1}{8Z_0Z_1}$

$P_{RISK} = \frac{12}{256}$

$P_{RISK} = 0.0468$

2.4 Part d

Given that $n_y = \sum_{i=1}^{n} 1(y_i = y)$

Using the distribution of the class-conditional densities, we get:

$\mu_0 = [0,0]^T$

$\mu_1 = [\frac{1}{n_1}\sqrt{2},0]^T$

$\Sigma_y = \sum_{i\in[1:n]:y_i=y} x_i - \mu_y$

$\Sigma_0 = \sum_{y=0} x_i$

$\Sigma_1 = \sum_{y=1} x_i - [\frac{\sqrt{2}}{n_1},0]^T$

2.5 Part e

$p(x|y=0) = N(\mu_0,\Sigma_0)$

$p(x|y=1) = N(\mu_1,\Sigma_1)$

$h_{QDA} = \arg\min[\frac{1}{2}(x-\mu_y)^T\Sigma_y^{-1}(x-\mu_y) + \frac{1}{2}\log\det(\Sigma_y) - \log(p(y))]$

$h_{QDA} = \arg\min[\frac{1}{2}||x-\mu_y||_{\Sigma_y} + \frac{1}{2}\log\det(\Sigma_y) - \log(\frac{1}{2})]$

$h_{QDA} = \arg\min[||x-\mu_y||_{\Sigma_y} + \log\det(\Sigma_y)]$

$P_{error} = \int_{R_2} L(x,y,h)p(x|y=0) + \int_{R_1} L(x,y,h)p(x|y=1)$

$P_{error} = \int_{R_2} (y-h(x))^2 N(x|\mu_0, \Sigma_0)dx + \int_{R_1} (y-h(x))^2 N(x|\mu_1, \Sigma_1)dx$

The error can be obtained by getting the squared loss of the true label versus the predicted label, over all overlapping regions and multiplying it by the Gaussian distribution in the region. To do this, we have to take the integral of the QDA rule as well as the Gaussian variable, evaluate the results over the two regions, and combine them for the total error.

2.6 Part f

$h_{LDA} - \arg\max[(\mu_y^T \Sigma^{-1})x - \frac{1}{2}\mu_y^T \Sigma^{-1}\mu_y + \log(p(y))]$

$w = \Sigma^{-1}(\mu_1 - \mu_0)$

$u = \frac{1}{2}(\mu_1 + \mu_0) - (\mu_1 - \mu_0)\frac{\log(p(y=1)) - \log(p(y=0))}{(\mu_1-\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0)}$

$h_{LDA} = 1(w^T(x-u) > 0)$

$h_{LDA} = 1(\Sigma^{-1}(\mu_1 - \mu_0)(x - \frac{1}{2}\mu_1 + \mu_0) > 0)$

$P_{error} = \int_{R_2} (y-h(x))^2 N(\mu_0, \Sigma) + \int_{R_1} (y-h(x))^2 N(\mu_1, \Sigma)$

$P_{error} = \int_{R_2} (y - 1(\Sigma^{-1}(\mu_1 - \mu_0)(x - \frac{1}{2}\mu_1 + \mu_0) > 0))^2 \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu_0)^2}{2\sigma^2}} + \int_{R_1}$

Since there is no loss if the labels match, we only care when $(y - 1(\Sigma^{-1}(\mu_1 - \mu_0)(x - \frac{1}{2}\mu_1 + \mu_0) > 0))^2 = 1$

$\int_R \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{2} erf(\frac{x-\mu}{\sigma\sqrt{2}})$

$P_{error} = \frac{1}{2} erf(\frac{x-\mu_0}{\sigma\sqrt{2}})|_{R_2} + \frac{1}{2} erf(\frac{x-\mu_1}{\sigma\sqrt{2}})|_{R_1}$

2.7 Part g

3 Comparing k-NN performance

3.1 Part a

$h_{MAP} = p(x|y)p(y)$

$h_{MAP} = \arg\max \frac{1}{2} Uniform(x|\mu_y + S)$

$P_{error} = \sum_{j=0}^{(k-1)/2} = \binom{k}{j} P(y|x)^j [1 - p(y|x)]^{k-j}$

3.2 Part b

3.3 Part c

3.4 Part d