University of California San Diego

ECE 271 Notes

Instructor: *Prof. Nuno Vasconcelos*

Organized by Ray Tsai

Bayes Decision Rule

$$
\begin{aligned}
g^*(x) &= \arg\min_{g(x)} \sum_i P_{Y|X}(i|x)L[g(x), i] \\
&= \arg\max_i P_{Y|X}(i|x) && \text{(for 0-1 loss function)} \\
&= \arg\max_i P_{X|Y}(x|i)P_Y(i) && \text{(for 0-1 loss function)} \\
&= \arg\max_i \log P_{X|Y}(x|i) + \log P_Y(i). && \text{(for 0-1 loss function)}
\end{aligned}
$$

For binary classification, the likelihood ratio form is: pick 0 if $\frac{P_{X|Y}(x|0)}{P_{X|Y}(x|1)} > T^* = \frac{P_Y(1)}{P_X(0)}$.

Associated Risk

$$
R^* = \int P_X(x) \sum_{i \neq g^*(x)} P_{Y|X}(i|x)dx = \int P_{Y,X}(y \neq g^*(x), x)dx \quad \text{(For 0-1 loss function)}
$$

Gaussian Classifier

For single variable, we assume $\sigma_i = \sigma$ and pick 0 if

$$
x < \frac{\mu_1 + \mu_0}{2} + \frac{1}{\frac{\mu_1 - \mu_0}{\sigma^2}} \log \frac{P_Y(0)}{P_Y(1)}.
$$

Generalizing it to multiple variables, we assume $\Sigma_i = \Sigma$, then the BDR becomes

$$
i^*(x) = \arg\min_i [d(x, \mu_i) + \alpha_i],
$$

where $d(x, y) = (x - y)^T \Sigma^{-1}(x - y)$ and $\alpha_i = \cancel{\log\left[(2\pi)^d|\Sigma|\right]} - 2\log P_Y(i)$.

Alternatively,

$$
i^*(x) = \arg\max_i g_i(x),
$$

where $g_i(x) = w_i^T x + w_{i0}$, $w_i = \Sigma^{-1}\mu_i$, and $w_{i0} = -\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \log P_Y(i)$.

Geometric Interpretation

Thus, the hyperplane between class 0 and 1 is

$$
g_0(x) - g_1(x) = w^T x + b = 0,
$$

where $w = \Sigma^{-1}(\mu_0 - \mu_1)$ and $b = -\frac{(\mu_0+\mu_1)^T \Sigma^{-1}(\mu_0-\mu_1)}{2} + \log \frac{P_Y(0)}{P_Y(1)}$.

It could also be rewritten as

$$
w^T(x - x_0) = 0,
$$

where $w = \Sigma^{-1}(\mu_0 - \mu_1)$ and $x_0 = \frac{\mu_0+\mu_1}{2} - \frac{1}{(\mu_0-\mu_1)^T \Sigma^{-1}(\mu_0-\mu_1)} \log \frac{P_Y(0)}{P_Y(1)}(\mu_0 - \mu_1)$

Gaussian Distribution Transformation

Let $x \sim N(\mu, \Sigma)$, and let $y = A^T x$, for some matrix A. Then, $y \sim N(A^T\mu, A^t\Sigma A)$. A special case of this is the whitening transform $A_w = \Phi\Lambda^{-1/2}$, where Φ is the matrix of orthonormal eigenvectors of Σ, and Λ is the diagonal matrix of eigenvalues of Σ.

Sigmoid

Suppose that $g_1(x) = 1 - g_0(x)$. Then, we can rewrite

$$g_0(x) = \frac{1}{1 + \frac{P_{X|Y}(x|1)P_Y(1)}{P_{X|Y}(x|0)P_Y(0)}} = \frac{1}{1 + \exp\{d_0(x, \mu_0) - d_1(x, \mu_1) + \alpha_0 - \alpha_1\}},$$

where, $d(x, y) = (x - y)^T \Sigma^{-1}(x - y)$ and $\alpha_i = \log\left[(2\pi)^d |\Sigma_i|\right] - 2\log P_Y(i)$.

Maximum Likelihood Estimation

Solve for

$$\Theta^* = \arg\max_{\Theta} P_X(D; \Theta) = \arg\max_{\Theta} \log P_X(D; \Theta).$$

Consider the Gaussian example:

Given a sample $\mathcal{D} = \{x_1, \dots, x_n\}$ of independent points, where $P_X(x_i) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)}$.

Then, the likelihood $L(x_1, \dots, x_n|\mu, \sigma) = \prod_{i=1}^n P_X(x_i)$. We take the gradient of the natrual log of L with respect to μ and get

$$\begin{aligned}
\nabla_\mu(\log L) &= \nabla_\mu \left(-\frac{1}{2}\log[(2\pi)^d|\Sigma|] - \frac{1}{2}\sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1}(x_i - \mu) \right) \\
&= \sum_{i=1}^n \Sigma^{-1}(x_i - \mu) = \sum_{i=1}^n x_i - \sum_{i=1}^n \mu = 0.
\end{aligned}$$

Thus, we get $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$. By taking the Hessian, we get $\nabla_\mu^2(\log L) = -\sum_{i=1}^n \Sigma^{-1} = -n\Sigma^{-1}$. Since the covariance matrix Σ is positive definite, $-n\Sigma^{-1}$ is negative definite. Thus $\hat{\mu}$ is the maximum point.

In addition, the MLE of the covariance matrix is

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T.$$

Bias and Variance

$$\begin{aligned}
Bias(\hat{\theta}) = E[\hat{\theta} - \theta], \quad Var(\hat{\theta}) = E\left\{(\hat{\theta} - E[\hat{\theta}])^2\right\}, \\
MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = Var(\hat{\theta}) + Bias^2(\hat{\theta}).
\end{aligned}$$

Least Squares

Consider a overdetermined system $\Phi\theta = z$, where we attempt to minimize $\|z - \Phi\theta\|$, the least square solution is

$$\theta^* = (\Phi^T\Phi)^{-1}\Phi^T z$$

For a overdetermined system $W\Phi\theta = Wz$, where we attempt to minimize $(z - \Phi\theta)^T W^T W(z - \Phi\theta)$, the least square solution is

$$\theta^* = (\Phi^T W^T W \Phi)^{-1}\Phi^T W^T W z.$$