Gaussian Discriminant Analysis

We assume that $p(x|y) \in \mathcal{N}(\mu, \Sigma)$ Multivariate Normal

1. Multivariate Normal Distribution

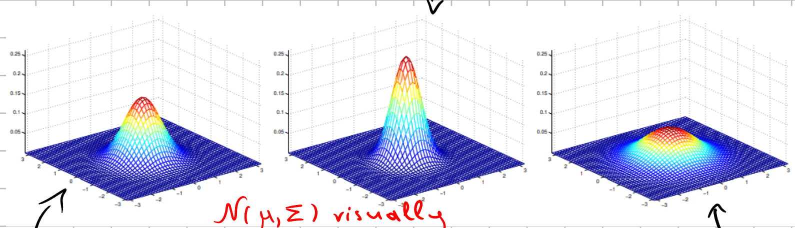
μ – mean vector $\in \mathbb{R}^d$, Σ – cov. matrix $\in \mathbb{R}^{d\times d}$, it is symmetric, positive semi-definite $\succeq 0$.

Its p.d.f. $p(x;\mu,\Sigma) = \dfrac{1}{2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

determinant of Σ.

also $\mu = 0$, but $\Sigma = 0{,}6I$

$\mu = \mathbb{E}(X) = \int_x x\, p(x;\mu,\Sigma)\, dx$



$\mathcal{N}(\mu,\Sigma)$ visually

$\mu = 0$, – 2x1 vector

$\Sigma = I$ – 2x2 Identity

Standard Normal Dist.

$\mu = 0, \Sigma = 2I$.

The GDA Model.

We have a classification problem, x – continuous-valued r.v.'s, we have a dataset $D = \{(x_i, y_i)\}_{i=1}^n$

$y \sim \text{Bernoulli}(\varphi)$

$x|y=0 \sim \mathcal{N}(\mu_0, \Sigma)$

$x|y=1 \sim \mathcal{N}(\mu_1, \Sigma)$

Both classes are normally distributed.

Thus, $p(y) = \varphi^y(1-\varphi)^{1-y}$; $p(x|y=0) = \dfrac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right)$;

μ_0

$p(x|y=1) = \dfrac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)$

μ_1

usually applied using one covariance matrix

Parameters of our model are $\varphi, \Sigma, \mu_0, \mu_1$

The likelihood of our model: $\mathcal{L} = \prod_{i=1}^n p(x_i, y_i; \varphi, \mu_0, \mu_1, \Sigma)$

The log-likelihood: $\ell = \sum_{i=1}^n \log\left(p(x_i, y_i; \varphi, \mu_0, \mu_1, \Sigma)\right) = \sum_{i=1}^n \log\left(p(x_i|y_i; \mu_0, \mu_1, \Sigma)\, p(y_i;\varphi)\right)$

We need to maximize it w.r.t. param-s.

Our objective: $\underset{\varphi, \mu_0, \mu_1, \Sigma}{\text{argmax}} \log\left\{\prod_{i=1}^n \left(p(x_i|y_i;\mu_0,\mu_1,\Sigma)\, p(y_i;\varphi)\right)\right\}$

1) $\left.\dfrac{\partial \ell}{\partial \varphi}\right|_{=0} = \dfrac{\partial}{\partial\varphi}\left(\log\left(\prod_{i=1}^n p(y_i;\varphi)\right)\right) = \dfrac{\partial}{\partial\varphi}\left(\log\left(\prod_{i=1}^n (\varphi^{y_i}(1-\varphi)^{1-y_i})\right)\right) = \dfrac{\partial}{\partial\varphi}\left(\sum_{i=1}^n \log(\varphi^{y_i}(1-\varphi)^{1-y_i})\right) = \dfrac{\partial}{\partial\varphi}\left(\sum_{i=1}^n \left(\log(\varphi^{y_i}) + (1-y_i)\log(1-\varphi)\right)\right) =$

здесь $\log = \ln$

$= \sum_{i=1}^n \left(\dfrac{\partial \log(\varphi)}{\partial\varphi}\cdot y_i + (1-y_i)\dfrac{\partial}{\partial\varphi}\left(\log(1-\varphi)\right)\right) = \sum_{i=1}^n \left(y_i \cdot \dfrac{1}{\varphi} - (1-y_i)\cdot\dfrac{1}{1-\varphi}\right) = \sum_{i=1}^n \left(\dfrac{y_i}{\varphi} - \dfrac{1-y_i}{1-\varphi}\right) = 0 \Rightarrow n_1 = \varphi(n_1+n_0) \Rightarrow \varphi^* = \dfrac{n_1}{n_1+n_0} = \dfrac{n_1}{n}$

Ну мы помним что $\dfrac{\partial}{\partial z}\ln(z) = \dfrac{1}{z}$

$\dfrac{\partial \ln(1-\varphi)}{\partial\varphi} = \dfrac{\partial(1-\varphi)}{\partial\varphi}\cdot\dfrac{1}{1-\varphi} = (-1)\dfrac{1}{1-\varphi}$

$\sum_{i=1}^n \left(y_i(1-\varphi) - (1-y_i)\varphi\right) = 0 \Rightarrow n_1(1-\varphi) - n_0\varphi = 0$

n объектов класса 1   n объектов класса 0

$n_1 - n_1\varphi - n_0\varphi = 0$

$n_1 - \varphi(n_1 + n_0)$

2) $\frac{\partial \ell}{\partial \mu_0} = \frac{\partial}{\partial \mu_0}\left\{ \log\left(\prod_{i=1}^{n_0} p(x_i \mid y_i = 0; \mu_0, \Sigma)\right)\right\} = \frac{\partial}{\partial \mu_0}\left\{ \log\left(\prod_{i=1}^{n_0}\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x_i-\mu_0)^T\Sigma^{-1}(x_i-\mu_0)\right]\right)\right)\right\} =$

$= \frac{\partial}{\partial \mu_0}\left\{ \sum_{i=1}^{n_0} \log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x_i-\mu_0)^T\Sigma^{-1}(x_i-\mu_0)\right]\right)\right\}$

$$\sum_{i=1}^{n_0}\left(\log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right) - \underbrace{\frac{1}{2}(x_i-\mu_0)^T\Sigma^{-1}(x_i-\mu_0)\, \overset{\ln e = 1}{\log e}}\right) = \sum_{i=1}^{n_0}\Sigma^{-1}(x_i-\mu_0) = 0 \;\Big/ \cdot \Sigma \implies \sum_{i=1}^{n_0}(x_i-\mu_0) = 0$$

$n_0 x_i - n_0\mu_0 = 0$

$\frac{\partial}{\partial \mu_0}\Big) = 0 \qquad \frac{\partial}{\partial \mu_0}\Big(= \frac{1}{2}\Sigma^{-1}\cdot 2(x_i-\mu_0)(-1) = -\Sigma^{-1}(x_i-\mu_0)$

$$\mu_0^* = \frac{n_0 x_i}{n_0}$$

3) Same procedure for $\mu_1 \implies \mu_1^* = \frac{n_1 x_i}{n_1}$

4) For Σ, $\frac{\partial \ell}{\partial \Sigma} = \frac{\partial}{\partial \Sigma}\left(\log\left(\prod_{i=1}^{n}\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x-\mu)\right]\right)\right) = \frac{\partial}{\partial \Sigma}\left(\sum_{i=1}^{n}\left(\log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right) +\right.\right.$

$\left.\left.+\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right)\right) = \frac{n}{2}\log|W| - \frac{1}{2}S_i^T = \frac{n}{2}\log W - \frac{1}{2}(x-\mu)(x-\mu)^T = 0$

$\frac{\partial}{\partial \Sigma} = 0$

$\log\left((2\pi)^{-d/2}|\Sigma|^{-\frac{1}{2}}\right) = \log(2\pi)\cdot\left(-\frac{d}{2}\right) - \frac{1}{2}\log(|\Sigma|) = \Big/ \sum_{i=1}^{n} \Big/ = -\frac{n}{2}\log|\Sigma| = \frac{n}{2}\log|W|$

$\sum_{i=1}^{n}\frac{\partial}{\partial \Sigma}\left(-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x-\mu)\right) = \sum_{i=1}^{n}\frac{\partial}{\partial \Sigma}\left(-\frac{1}{2}tr((x-\mu)(x-\mu)^T\Sigma^{-1})\right) = \Big/$ let's differentiate w.r.t. $\Sigma^{-1} = W \Big/ =$

$\log W = -\log \Sigma$

$= \frac{\partial}{\partial W}\left(\sum_{i=1}^{n} -\frac{1}{2}tr((x-\mu)(x-\mu)^T W)\right) = -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial tr((x-\mu)(x-\mu)^T W)}{\partial W} = -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial tr(S_i W)}{\partial W} = -\frac{1}{2}S_i^T$

let $(x-\mu)(x-\mu)^T = S_i$

$\frac{n}{2}\log|\Sigma| = \frac{1}{2}\sum_{i=1}^{n}(x-\mu)(x-\mu)^T \implies \Sigma^* = \frac{1}{n}\sum_{i=1}^{n}(x-\mu)(x-\mu)^T$

The optimal values are:

$$\varphi = \frac{1}{n}\sum_{i=1}^{n} 1\{y_i = 1\}; \quad \mu_0 = \frac{\sum_{i=1}^{n} 1\{y_i=0\}x_i}{\sum_{i=1}^{n} 1\{y_i = 0\}}; \quad \mu_1 = \frac{\sum_{i=1}^{n} 1\{y_i=1\}x_i}{\sum_{i=1}^{n} 1\{y_i=1\}}; \quad \Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_{y_i})(x_i-\mu_{y_i})^T$$

Two Gaussians have contours that are the same shape and orientation, since they share a cov. matrix Σ.