5/17/2019 Discussion # 7

## The multivariate normal distribution

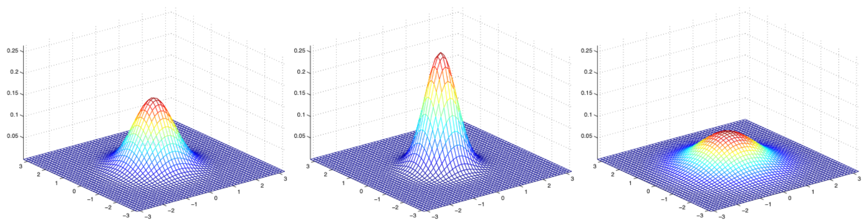
$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

$\underline{\mu} \in R^{n\times n}$, $\underline{\underline{\Sigma}} \in R^{n\times n}$ $|\underline{\underline{\Sigma}}|$ determinant.

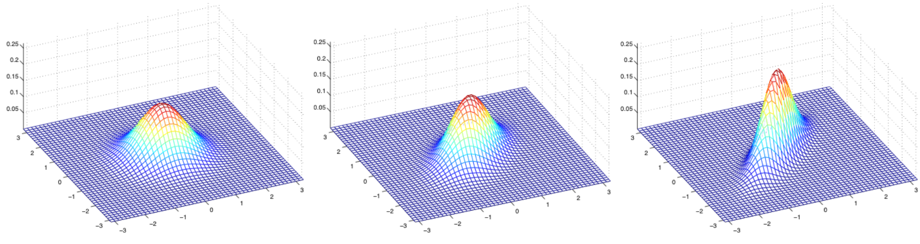
$$\mathrm{E}[X] = \int_x x\, p(x;\mu,\Sigma)dx = \mu$$

$$\mathrm{Cov}(X) = \Sigma.$$

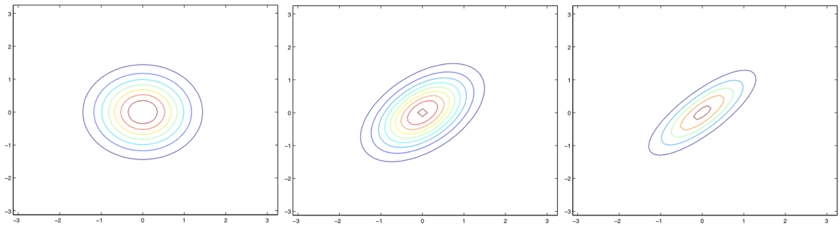
Fixed $\underline{\underline{u}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$



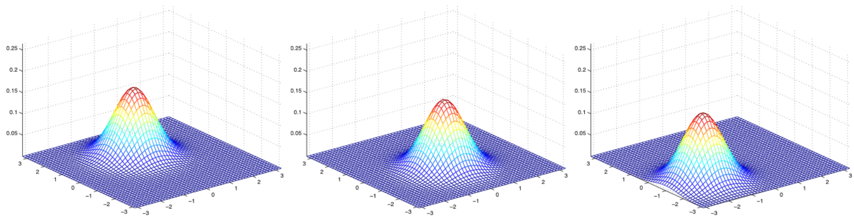

$\underline{\underline{\Sigma}} = \underline{\underline{I}}$ $\underline{\underline{\Sigma}} = 0.6\underline{\underline{I}}$ $\Sigma = 2I$



$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}; \quad \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$



Effect of $\mu$, Fixed $\underline{\underline{\Sigma}} = \underline{\underline{I}}$



$\underline{\underline{\Sigma}} = \underline{\underline{I}}$

$$\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \quad \mu = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix}; \quad \mu = \begin{bmatrix} -1 \\ -1.5 \end{bmatrix}$$

# The Gaussian Discriminant Analysis model

$$
\begin{aligned}
y &\sim \text{Bernoulli}(\phi) \\
x|y=0 &\sim \mathcal{N}(\mu_0, \Sigma) \\
x|y=1 &\sim \mathcal{N}(\mu_1, \Sigma)
\end{aligned}
$$

$y \in \{0, 1\}.$

$$
\begin{aligned}
p(y) &= \underline{\phi^y(1-\phi)^{1-y}} \\
p(x|y=0) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\underline{\mu_0})^T\underline{\Sigma}^{-1}(x-\mu_0)\right) \\
p(x|y=1) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\underline{\mu_1})^T\underline{\Sigma}^{-1}(x-\mu_1)\right)
\end{aligned}
$$

$$
\begin{cases} p(y=1) = \phi \\ p(y=0) = 1-\phi \end{cases}
$$

**training set:**

$(\underline{x}_1, y_1), \ldots (\underline{x}_m, y_m)$

$$\underset{\phi, \mu_0, \mu_1, \Sigma}{\arg\max} \; p(\underline{x}_1, \ldots \underline{x}_m, y_1, \ldots y_m)$$

Defined:

$$L(\phi, \mu_0, \mu_1, \Sigma) = \log p(x_1, \ldots x_m, y_1, \ldots y_m)$$

Convert original problem to

$$\underset{\phi, \mu_0, \mu_1, \Sigma}{\arg\max} \; L(\phi, \mu_0, \mu_1, \Sigma)$$

$$p(\underline{x}_1, \dots \underline{x}_m, y_1, \dots, y_m)$$

$$= \prod_{i=1}^{m} p(x_i, y_i)$$

$$= \prod_{i=1}^{m} p(x_i | y_i)\, p(y_i)$$

$$\begin{aligned} p(y) &= \phi^y (1-\phi)^{1-y} \\ p(x|y=0) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)\right) \\ p(x|y=1) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right) \end{aligned}$$

$p(y_i=0)$

$$= \prod_{i=1}^{m} \left[\frac{1-\phi}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1} (x-\mu_0)\right)\right]^{1-y_i}$$

$p(y_i=1)$

$$\times \left[\frac{\phi}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)\right)\right]^{y_i}$$

$$L(\phi, \mu_0, \mu_1, \Sigma) = \log p(\underline{x}_1, \dots \underline{x}_m, y_1, \dots, y_m)$$

$$L(\phi, \mu_0, \mu_1, \Sigma) = \log p(\underline{x}_1, \dots \underline{x}_m, y_1, \dots, y_m)$$

$$= \sum_{i=1}^{m} \left\{ \log \left[ \frac{1-\phi}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1} (x-\mu_0)\right) \right]^{1-y_i} + \log \left[ \frac{\phi}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1)\right) \right]^{y_i} \right\}$$

$$= \sum_{i=1}^{m} \left\{ (1-y_i) \left[ \ln(1-\phi) - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2}(x-\mu_0)^T \Sigma^{-1} (x-\mu_0) \right] + y_i \left[ \ln \phi - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma| - \frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1) \right] \right\}$$

$$\underset{\phi, \mu_1, \mu_2, \Sigma}{\arg\max} \; L(\phi, \mu_0, \mu_1, \Sigma) \leftarrow \text{concave function}$$

take partial derivative of $L$, and set it to zero, find:

$$\phi = \frac{1}{m} \sum_{i=1}^{m} 1\{y^{(i)} = 1\}$$

$$\mu_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\}}$$

$$\mu_1 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}$$

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T.$$

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

$$\frac{\partial}{\partial \Sigma}\Bigg\{\sum_{i=1}^{m}\Big\{(1-y_i)\big[\ln(1-\phi) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(x_i-\mu_0)^T\Sigma^{-1}(x_i-\mu_0)\big]$$

$$+ \; y_i \; \big[\ln(\phi) - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(x_i-\mu_1)^T\Sigma^{-1}(x_i-\mu_1)\big]\Big\}\Bigg\} \quad (1)$$

Aside:

1. if $a \in \mathbb{R}$, $a = Tr(a)$
2. $Tr(A+B) = Tr(A) + Tr(B)$
3. $\frac{\partial \ln|A|}{\partial A} = A^{-T}$
4. $\frac{\partial Tr(A^{-1}B)}{\partial A} = -(A^{-1}BA^{-1})^T$
5. $Tr(AB) = Tr(BA)$
6. $Tr(ABC) = Tr(CBA)$

$$\frac{\partial}{\partial \Sigma} L = \frac{\partial}{\partial \Sigma}\Big\{-\frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m}(x_i-\mu_{y^{(i)}})^T\Sigma^{-1}(x_i-\mu_{y_i})\Big\}$$

$$= \frac{\partial}{\partial \Sigma}\Big\{-\frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m} Tr\big[(x_i-\mu_{y^{(i)}})^T\Sigma^{-1}(x_i-\mu_{y_i})\big]\Big\}$$

aside at the end

$$= \frac{\partial}{\partial \Sigma}\Big\{-\frac{m}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{m} Tr\big[\Sigma^{-1}(x_i-\mu_{y^{(i)}})(x_i-\mu_{y_i})^T\big]\Big\}$$

$$= \frac{\partial}{\partial \Sigma}\Big\{-\frac{m}{2}\ln|\Sigma| - \frac{m}{2} Tr\Big(\Sigma^{-1}\underbrace{\frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_{y_i})(x_i-\mu_{y_i})^T}_{S}\Big)$$

$$= \frac{\partial}{\partial \Sigma}\Big(-\frac{m}{2}\ln|\Sigma| - \frac{m}{2} Tr(\Sigma^{-1}S)\Big)$$

$$\frac{\partial}{\partial \Sigma}\left(-\frac{m}{2}\ln|\Sigma| - \frac{m}{2}Tr(\Sigma^{-1}S)\right)$$

Aside:

1. $\frac{\partial \ln|A|}{\partial A} = A^{-T}$ 2. $\frac{\partial Tr(A^{-1}B)}{\partial A} = -(A^{-1}BA^{-1})^T$

$$-\frac{m}{2}\Sigma^{-T} + \frac{m}{2}(\Sigma^{-1}S\Sigma^{-1})^T = 0$$

$$\Rightarrow \Sigma^{-T} = (\Sigma^{-1}S\Sigma^{-1})^T$$

$$\Rightarrow \Sigma^{-1} = (\Sigma^{-1}S\Sigma^{-1})$$

$$\Sigma\Sigma^{-1}\Sigma = \Sigma(\Sigma^{-1}S\Sigma^{-1})\Sigma$$

$$\Sigma = S$$

Aside:

$$\frac{1}{2}\sum_{i=1}^{m} Tr[\Sigma^{-1}(x_i - \mu_{y_i})(x_i - \mu_{y_i})^T]$$

$$= \frac{m}{2}\sum_{i=1}^{m}\frac{1}{m} Tr[\Sigma^{-1}(x_i - \mu_{y_i})(x_i - \mu_{y_i})^T)$$

$\downarrow$ $a\,Tr(\underline{A}) = Tr(a\underline{A})$

$$= \frac{m}{2}\sum_{i=1}^{m} Tr[\frac{1}{N}\Sigma^{-1}(x_i - \mu_{y_i})(x_i - \mu_{y_i})^T)$$

$\downarrow$ $Tr(\underline{A}+\underline{B}) = Tr(\underline{A}) + Tr(\underline{B})$

$$= \frac{m}{2} Tr[\sum_{i=1}^{m}\Sigma^{-1}\frac{1}{m}(x_i - \mu_{y_i})(x_i - \mu_{y_i})^T)$$

$$= \frac{m}{2} Tr[\Sigma^{-1}\frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{y_i})(x_i - \mu_{y_i})^T)$$

# problem 1 Discussion 7

(1) Logistic Regression

$\sigma(\underline{w}^T \underline{x} + b)$

parameters: $\underline{w}$, $b$

$\underline{w} \in R^m$ needs: $\boxed{m+1}$ to determine $\underline{w}$ and $b$

$\underline{b} \in R^1$

(2) GDA:

square and symmetry

parameters: $\underset{1}{\phi}$, $\underset{m}{\underline{\mu}_0}$, $\underset{m}{\underline{\mu}_1}$, $\underset{\frac{m(m+1)}{2}}{\underline{\underline{\Sigma}}}$

need $1 \quad m \quad m \quad \frac{m(m+1)}{2}$

total: $\boxed{1+2m+\frac{m(m+1)}{2}}$

e.g. 3x3 symmetry matrix.

$$\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

only need a, b, c, d, e f totally 6 parameters to fully determine a 3x3 symmetric matrix.

$1+2+3 = \frac{3(3+1)}{2} = 6.$

(3) Navie Bays:

diagnal matrix.

parameters: $\phi$, $p(x_i|y=0)$, $p(x_i|y=1)$, $\underline{\underline{\sigma}}^2$

$1 \quad m \quad m \quad m$

total: $\boxed{1+3m}$