UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escola Tècnica Superior d'Enginyeria Informàtica

etsinf

Tema 7. Distribución Gaussiana

Percepción (PER)

Curso 2017/2018

Departamento de Sistemas Informáticos y Computación

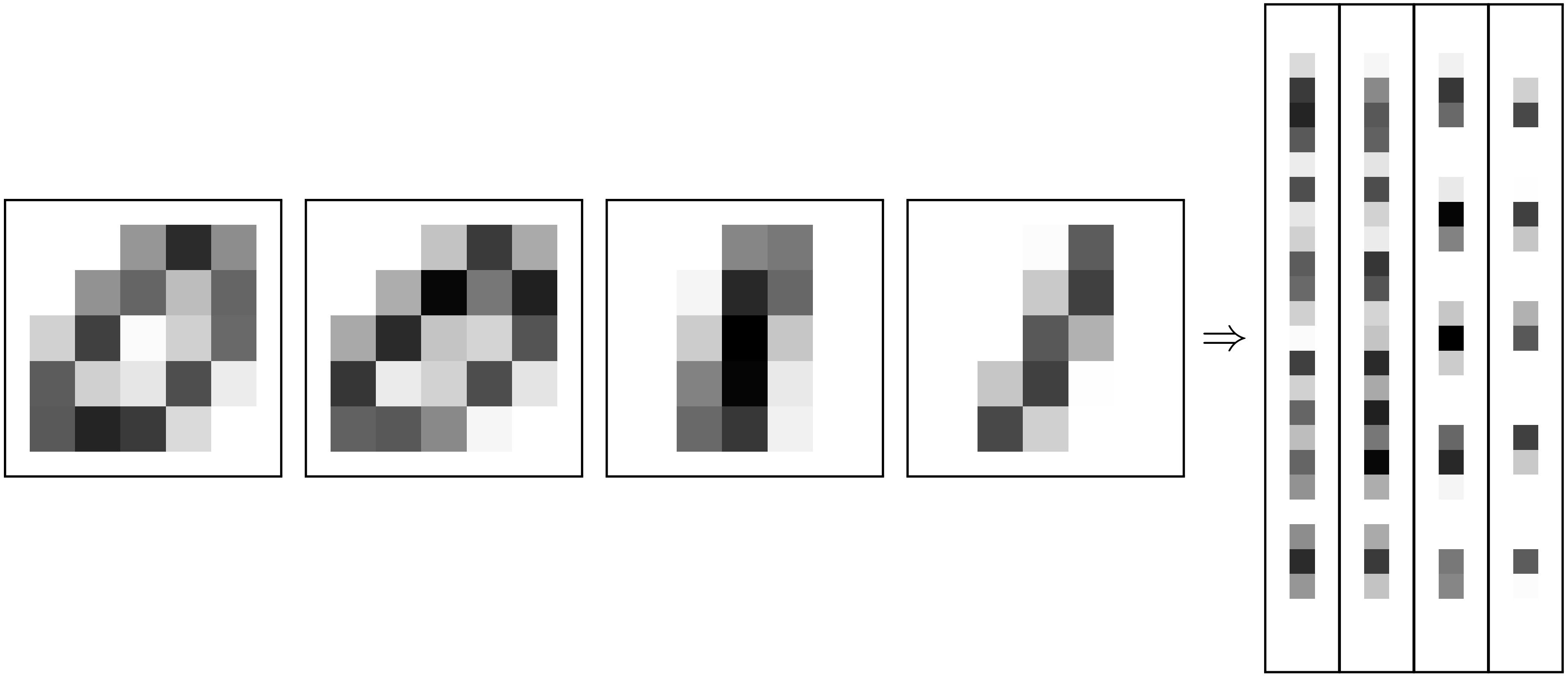
Índice

Índice

Introducción y motivación

Algunas tareas representan objetos por *vectores de características reales* ($\mathbb{R}^D$)

Ejemplo: Imágenes de 5×5 en escala de grises interpretadas como vectores reales de características de 25 dimensiones.



Ejemplo: Señal acústica mediante vectores de coeficientes cepstrales

Idea: usar la *distribución gaussiana* para modelizar la condicional $p(\boldsymbol{x}|c)$

Índice

Definición: distribución gaussiana unidimensional

Sea x una variable aleatoria unidimensional

Gaussiana unidimensional estandarizada

$x \sim N(0, 1)$ presenta una distribución de probabilidad

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\,x^2\right)$$

Gaussiana unidimensional general

$x \sim N(\mu, \sigma)$, con media $\mu \in \mathbb{R}$ y varianza $\sigma^2 \in \mathbb{R}^+$, presenta una distribución de probabilidad

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Definición: distribución gaussiana multidimensional

Sea $\boldsymbol{x} = (x_1, \ldots, x_D)^t$ una variable aleatoria D-dimensional

Gaussiana estandarizada

$\boldsymbol{x} \sim N_D(\mathbf{0}, I_D)$, donde $x_1, \ldots, x_D \sim N(0, 1)$ independientes, presenta una distribución de probabilidad:

$$p(\boldsymbol{x}) = (2\pi)^{-\frac{D}{2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{x}\right)$$

Definición: distribución gaussiana multidimensional

Gaussiana general

Sean:

- $\boldsymbol{z} \sim N_D(\mathbf{0}, I_D)$
- $\boldsymbol{\mu} \in \mathbb{R}^D$
- $A \in \mathbb{R}^{D \times D} : |A| \neq 0$
- $\Sigma = AA^t$ (simétrica y definida positiva) con:
  - $A = W\Delta^{\frac{1}{2}}$
  - W vectores propios de Σ
  - Δ valores propios de Σ
- $\boldsymbol{x} = A\boldsymbol{z} + \boldsymbol{\mu}$

$\boldsymbol{x} \sim N_D(\boldsymbol{\mu}, \Sigma)$, con media $\boldsymbol{\mu}$ y matriz de covarianzas Σ, presenta una distribución de probabilidad:

$$p(\boldsymbol{x}) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

Índice

Clasificador gaussiano

Clasificador gaussiano: clasificador de Bayes donde la f.d. condicional $p(\boldsymbol{x}|c)$ es una gaussiana

$$p(\boldsymbol{x} \mid c) \sim N_D(\boldsymbol{\mu}_c, \Sigma_c), \quad c = 1, \ldots, C.$$

Por tanto:

$$\begin{aligned}
c^*(\boldsymbol{x}) &= \underset{c=1,\ldots,C}{\text{argmax}} \ \log\ P(c) + \log\ p(\boldsymbol{x} \mid c) \\
&= \underset{c=1,\ldots,C}{\text{argmax}} \ \log\ P(c) - \frac{D}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_c| - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_c)^t \Sigma_c^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_c) \\
&= \underset{c=1,\ldots,C}{\text{argmax}} \ \log P(c) - \frac{1}{2}\log|\Sigma_c| - \frac{1}{2}\boldsymbol{x}^t \Sigma_c^{-1} \boldsymbol{x} + \boldsymbol{\mu}_c^t \Sigma_c^{-1} \boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_c^t \Sigma_c^{-1} \boldsymbol{\mu}_c \\
&= \underset{c=1,\ldots,C}{\text{argmax}} \ -\frac{1}{2}\boldsymbol{x}^t \Sigma_c^{-1} \boldsymbol{x} + \boldsymbol{\mu}_c^t \Sigma_c^{-1} \boldsymbol{x} + \left(\log P(c) - \frac{1}{2}\log|\Sigma_c| - \frac{1}{2}\boldsymbol{\mu}_c^t \Sigma_c^{-1} \boldsymbol{\mu}_c \right)
\end{aligned}$$

Clasificador gaussiano

$$c^*(\boldsymbol{x}) = \operatorname*{argmax}_{c=1,\ldots,C} \; -\frac{1}{2}\boldsymbol{x}^t\Sigma_c^{-1}\boldsymbol{x}+\boldsymbol{\mu}_c^t\Sigma_c^{-1}\boldsymbol{x}+\left(\log P(c) - \frac{1}{2}\log|\Sigma_c| - \frac{1}{2}\boldsymbol{\mu}_c^t\Sigma_c^{-1}\boldsymbol{\mu}_c\right)$$

Clasificador ***cuadrático*** con $\boldsymbol{x}$:

$$c^*(\boldsymbol{x}) = \operatorname*{argmax}_{c=1,\ldots,C} \; g_c(\boldsymbol{x}) = \operatorname*{argmax}_{c=1,\ldots,C} \boldsymbol{x}^t\, W_c\, \boldsymbol{x} + \boldsymbol{w}_c^t\, \boldsymbol{x} + w_{c0}$$

Con:

$$W_c = -\frac{1}{2}\Sigma_c^{-1} \qquad \boldsymbol{w}_c = \Sigma_c^{-1}\boldsymbol{\mu}_c$$

$$w_{c0} = \log P(c) - \frac{1}{2}\log|\Sigma_c| - \frac{1}{2}\boldsymbol{\mu}_c^t\Sigma_c^{-1}\boldsymbol{\mu}_c$$

Clasificador gaussiano

Caso particular: ***matriz de covarianzas común***, $\Sigma_c = \Sigma$

En ese caso, tanto $-\frac{1}{2}\boldsymbol{x}^t\Sigma^{-1}\boldsymbol{x}$ como $-\frac{1}{2}\log|\Sigma|$ son independientes de c

$$c^*(\boldsymbol{x}) = \underset{c=1,\dots,C}{\text{argmax}} \; \boldsymbol{\mu}_c^t\Sigma^{-1}\boldsymbol{x} + \left(\log P(c) - \frac{1}{2}\boldsymbol{\mu}_c^t\Sigma^{-1}\boldsymbol{\mu}_c\right)$$

El clasificador gaussiano es ***lineal***:

$$c^*(\boldsymbol{x}) = \underset{c=1,\dots,C}{\text{argmax}} \; g_c(\boldsymbol{x}) = \underset{c=1,\dots,C}{\text{argmax}} \, \boldsymbol{w}_c^t\boldsymbol{x} + w_{c0}$$

Con:

$$\boldsymbol{w}_c = \Sigma^{-1}\boldsymbol{\mu}_c \qquad w_{c0} = \log P(c) - \frac{1}{2}\boldsymbol{\mu}_c^t\Sigma^{-1}\boldsymbol{\mu}_c$$

Índice

Entrenamiento por máxima verosimilitud

Sean N muestras de entrenamiento aleatoriamente extraídas de C distribuciones gaussianas independientes

$$\{(\boldsymbol{x}_n, c_n)\}_{n=1}^{N} \quad \text{i.i.d.} \quad p(\boldsymbol{x}, c) = P(c)\, p(\boldsymbol{x}|c), \quad p(\boldsymbol{x}|c) \sim N_D(\boldsymbol{\mu}_c, \Sigma_c)$$

Conjunto de parámetros a estimar Θ:

- Probabilidades *a priori*: $P(1), \dots, P(C)$
- Medias para cada clase: $\mu_1, \dots, \mu_C$
- Matrices de covarianza para cada clase: $\Sigma_1, \dots, \Sigma_C$

Por ***criterio de máxima verosimilitud*** (MV), se estima $\boldsymbol{\Theta}$ como:

$$\hat{P}(c) = \frac{N_c}{N} \qquad (1) \qquad\qquad \hat{\boldsymbol{\mu}}_c = \frac{1}{N_c} \sum_{n:c_n=c} \boldsymbol{x}_n \qquad (2)$$

$$\hat{\Sigma}_c = \frac{1}{N_c} \sum_{n\,:\,c_n=c} (\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_c)(\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_c)^t = \left(\frac{1}{N_c} \sum_{n\,:\,c_n=c} \boldsymbol{x}_n \boldsymbol{x}_n^t \right) - \hat{\boldsymbol{\mu}}_c \hat{\boldsymbol{\mu}}_c^t \qquad (3)$$

Entrenamiento por máxima verosimilitud

En el caso de Σ común para todas las clases ($\Sigma_c = \Sigma$), el conjunto de parámetros a estimar Θ es:

- Probabilidades *a priori*: $P(1), \ldots, P(C)$
- Medias para cada clase: $\mu_1, \ldots, \mu_C$
- Matriz de covarianza común: Σ

Por ***criterio de máxima verosimilitud***, la estimación de $\boldsymbol{\Theta}$ se calcula como en el caso general (Ecuaciones (1) y (2) para $\hat{P}(c)$ y $\hat{\boldsymbol{\mu}}_c$, respectivamente) y :

$$\hat{\Sigma} = \sum_c \hat{P}(c)\,\hat{\Sigma}_c = \frac{1}{N}\sum_n \boldsymbol{x}_n \boldsymbol{x}_n^t - \sum_c \hat{P}(c)\,\hat{\boldsymbol{\mu}}_c \hat{\boldsymbol{\mu}}_c^t \qquad (4)$$

Con $\hat{\Sigma}_c$ calculada según Ecuación (3).

Índice

Suavizado

Umbralizado de covarianza [Duda01, pág. 113]

Covarianzas con magnitud de la correlación no cercana a uno valen cero:

$$\tilde{\sigma}^2_{cdd'} = \begin{cases} \hat{\sigma}^2_{cdd'} & \text{si } |\hat{\rho}_{cdd'}| > 1 - \epsilon \\ 0 & \text{otro caso} \end{cases} \qquad \forall c, d, d' = 1, \dots, D; \; d \neq d'$$

Donde:

- ϵ es una constante pequeña no negativa ($\epsilon = 0 \rightarrow \Sigma$ diagonal)
- Coeficiente de correlación: $\hat{\rho}_{cdd'} = \frac{\hat{\sigma}^2_{cdd'}}{\hat{\sigma}_{cdd}\,\hat{\sigma}_{cd'd'}}$

Flat smoothing

Combinación lineal de cada $\hat{\Sigma}_c$ y $\tilde{\Sigma}$ (matriz de covarianza global suavizada):

$$\tilde{\Sigma}_c = \alpha\,\hat{\Sigma}_c + (1 - \alpha)\,\tilde{\Sigma} \qquad \forall c \quad \alpha \in [0, 1]$$

Donde: $\tilde{\Sigma} = \beta\hat{\Sigma} + (1 - \beta)I$, $\beta \in [0, 1]$