Elementy Statystycznego Uczenia Maszynowego

Modele Gaussowskie

Wielowymiarowy rozkład normalny

Powtórka z rachunku prawdopodobieństwa:

- Jeśli $\mathbf{x} = [x_1, x_2, \ldots, x_d]^\mathsf{T}$ ma wielowymiarowy rozkład normalny (*Multivariate Normal distribution* – MVN) z wartością oczekiwaną $\boldsymbol{\mu}$ i macierzą kowariancji $\boldsymbol{\Sigma}$ to:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$$

  Co możemy wyrazić skrótowo:

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Ponadto:

$$\boldsymbol{\Lambda} := \boldsymbol{\Sigma}^{-1} \text{ - macierz precyzji.}$$

Wnioskowanie w łącznym rozkładzie MVN

Niech

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_l \end{bmatrix}$$

mają łącznie rozkład normalny:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

z parametrami:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{xx}} & \boldsymbol{\Sigma}_{\mathbf{xy}} \\ \boldsymbol{\Sigma}_{\mathbf{yx}} & \boldsymbol{\Sigma}_{\mathbf{yy}} \end{bmatrix}, \ \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{xx}} & \boldsymbol{\Lambda}_{\mathbf{xy}} \\ \boldsymbol{\Lambda}_{\mathbf{yx}} & \boldsymbol{\Lambda}_{\mathbf{yy}} \end{bmatrix}$$

Wnioskowanie w łącznym rozkładzie MVN

Wówczas:

- Rozkłady brzegowe są rozkładami normalnymi:

$$\mathbf{x} \sim N\left(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{xx}}\right),$$

$$\mathbf{y} \sim N\left(\boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\Sigma}_{\mathbf{yy}}\right)$$

Wnioskowanie w łącznym rozkładzie MVN

Wówczas:

- Rozkłady warunkowe są rozkładami normalnymi:

$$\mathbf{x} \mid \mathbf{y} \sim N\left(\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}, \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}}\right),$$

gdzie:

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}} &= \boldsymbol{\Sigma}_{\mathbf{xx}} - \boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}{}^{-1}\boldsymbol{\Sigma}_{\mathbf{yx}} \\
&= \boldsymbol{\Lambda}_{\mathbf{xx}}{}^{-1}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} &= \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}{}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right) \\
&= \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}}\left(\boldsymbol{\Lambda}_{\mathbf{xx}}\boldsymbol{\mu}_{\mathbf{x}} - \boldsymbol{\Lambda}_{\mathbf{xy}}\left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}\right)\right)
\end{aligned}$$

Liniowe modele Gaussowskie

Powyższe własności rozkładów łącznych pozwalają wnioskować (jawnie) w *linowych modelach Gaussowskich* (*Linear Gaussian Models*).

Niech $\mathbf{x} = [x_1, x_2, \ldots, x_k]^\mathsf{T}$ będzie pewną zmienną ukrytą (nieobserwowaną), zaś:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} + \boldsymbol{\epsilon}$$

jej pomiarem. Zakładamy, że pomiar dokonany jest ze (znaną) precyzją ${\boldsymbol{\Sigma}_\mathbf{y}}^{-1}$:

$$\boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{y}\right)$$

Ponadto zakładamy, że macierz $\mathbf{A}_{l \times k}$ oraz wektor $\mathbf{b}_{l \times 1}$ są znane.

Liniowe modele Gaussowskie

Przy tych założeniach wiarygodność ma postać:

$$p\left(\mathbf{y} \mid \mathbf{x}\right) = N\left(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{\Sigma}_{\mathbf{y}}\right)$$

Dalej, przyjmijmy rozkład normalny jako rozkład a priori dla $\mathbf{x}$ (w tym przypadku jest to rozkład sprzężony do wiarygodności):

$$\mathbf{x} \sim N\left(\mu_{\mathbf{x}}, \mathbf{\Sigma}_{\mathbf{x}}\right)$$

Pytanie brzmi: jaką postać ma rozkład a posteriori $p\left(\mathbf{x} \mid \mathbf{y}\right)$?

Jest to klasyczny przykład liniowego modelu Gaussowskiego.

Liniowe modele Gaussowskie

Formalnie, mamy model postaci:

$$\mathbf{x} \sim N\left(\mu_{\mathbf{x}}, \mathbf{\Sigma}_{\mathbf{x}}\right)$$

$$\mathbf{y} \mid \mathbf{x} \sim N\left(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{\Sigma}_{\mathbf{y}}\right)$$

Wówczas (korzystając z rozkładu warunkowego w MVN) otrzymujemy rozkład a posteriori postaci

$$\mathbf{x} \mid \mathbf{y} \sim N\left(\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}, \mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}}\right)$$

gdzie:

$$\mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}} = \left[\mathbf{\Sigma}_{\mathbf{x}}{}^{-1} + \mathbf{A}^{\mathsf{T}}\mathbf{\Sigma}_{\mathbf{y}}{}^{-1}\mathbf{A}\right]^{-1}$$

$$\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}}\left[\mathbf{A}^{\mathsf{T}}\mathbf{\Sigma}_{\mathbf{y}}{}^{-1}(\mathbf{y} - \mathbf{b}) + \mathbf{\Sigma}_{\mathbf{x}}{}^{-1}\boldsymbol{\mu}_{\mathbf{x}}\right]$$