Wrocław University of Science and Technology

Regresja liniowa

Szymon Zaręba

szymon.zareba@pwr.edu.pl

Probabilistyczny model regresji liniowej

Model

Przyjmijmy następujący model:

$$y = \boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{w} + \varepsilon,$$

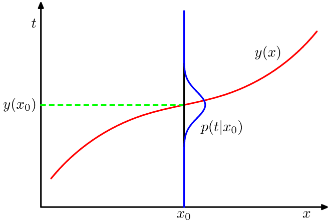
gdzie $\varepsilon \sim \mathcal{N}(\varepsilon|0, \sigma^2)$ oznacza zakłócenie. Zauważmy, że

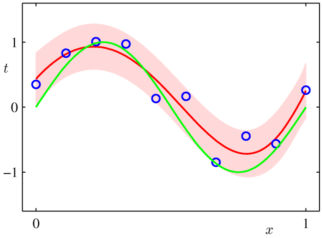
$$p(y|x, \mathbf{w}, \sigma^2) = \mathcal{N}(y|\boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{w}, \sigma^2),$$

ponieważ

$$\mathbb{E}[y] = \mathbb{E}[\boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{w} + \varepsilon] = \boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{w}$$

$$\mathrm{Var}[y] = \mathrm{Var}[\boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{w} + \varepsilon] = \sigma^2$$





Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć funkcję wiarygodności:

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć funkcję wiarygodności:

$$p(Y|X, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n|x_n, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(y_n|\boldsymbol{\phi}(x_n)^{\mathrm{T}}\mathbf{w}, \sigma^2),$$

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć funkcję wiarygodności:

$$p(Y|X, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n|x_n, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(y_n|\boldsymbol{\phi}(x_n)^{\mathrm{T}}\mathbf{w}, \sigma^2),$$

po wstawieniu rozkładu normalnego:

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć funkcję wiarygodności:

$$p(Y|X, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} p(y_n|x_n, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(y_n|\boldsymbol{\phi}(x_n)^{\mathrm{T}}\mathbf{w}, \sigma^2),$$

po wstawieniu rozkładu normalnego:

$$p(Y|X, \mathbf{w}, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y_n - \boldsymbol{\phi}(x_n)^{\mathrm{T}}\mathbf{w})^2}{2\sigma^2}$$

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć logarytm funkcji wiarygodności:

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć logarytm funkcji wiarygodności:

$$\ln p(Y|X, \mathbf{w}, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}||Y - \Phi\mathbf{w}||_2^2$$

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć logarytm funkcji wiarygodności:

$$\ln p(Y|X, \mathbf{w}, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}||Y - \Phi\mathbf{w}||_2^2$$

wyznaczyć jego gradient:

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć logarytm funkcji wiarygodności:

$$\ln p(Y|X, \mathbf{w}, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}||Y - \Phi\mathbf{w}||_2^2$$

wyznaczyć jego gradient:

$$\nabla_w \ln p(Y|X, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2}(-2\Phi^\intercal Y + 2\Phi^\intercal\Phi\mathbf{w})$$

Probabilistyczny model regresji liniowej

Uczenie

Można wyznaczyć logarytm funkcji wiarygodności:

$$\ln p(Y|X, \mathbf{w}, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}||Y - \Phi\mathbf{w}||_2^2$$

wyznaczyć jego gradient:

$$\nabla_w \ln p(Y|X, \mathbf{w}, \sigma^2) = -\frac{1}{2\sigma^2}(-2\Phi^\intercal Y + 2\Phi^\intercal\Phi\mathbf{w})$$

i rozwiązać względem $\mathbf{w}$:

$$-2\Phi^\intercal Y + 2\Phi^\intercal\Phi\mathbf{w} = 0$$

$$\mathbf{w} = (\Phi^\intercal\Phi)^{-1}\Phi^\intercal Y$$

Probabilistyczny model regresji liniowej

Predykcja

Teoria decyzji mówi, że optymalnym wyborem jest y, który minimalizuje ryzyko (średnią stratę) przy zadanej funkcji straty. Mając:

- $p(y|x, \mathbf{w})$ - model
- $y \in \mathbb{R}$ - zbiór możliwych decyzji
- $L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$ - funkcja straty

$$R(\hat{y}) = \mathbb{E}[L(y, \hat{y})] = \int \frac{1}{2}(y - \hat{y})^2 p(y|x) dy$$

$$y^* = \operatorname*{argmin}_{\hat{y}} R(\hat{y})$$

Probabilistyczny model regresji liniowej

Predykcja

Należy policzyć pochodną ryzyka względem podejmowanej decyzji:

Probabilistyczny model regresji liniowej

Predykcja

Należy policzyć pochodną ryzyka względem podejmowanej decyzji:

$$\frac{\partial R(\hat{y})}{\partial \hat{y}} = -\int (y - \hat{y}) p(y|x) dy$$

Probabilistyczny model regresji liniowej

Predykcja

Należy policzyć pochodną ryzyka względem podejmowanej decyzji:

$$\frac{\partial R(\hat{y})}{\partial \hat{y}} = -\int (y - \hat{y}) p(y|x) dy$$

i wyznaczyć optymalną wartość decyzji:

Probabilistyczny model regresji liniowej

Predykcja

Należy policzyć pochodną ryzyka względem podejmowanej decyzji:

$$\frac{\partial R(\hat{y})}{\partial \hat{y}} = -\int (y - \hat{y}) p(y|x) dy$$

i wyznaczyć optymalną wartość decyzji:

$$-\int (y - \hat{y}) p(y|x) dy = 0$$

$$\int y p(y|x) dy - \int \hat{y} p(y|x) dy = 0$$

$$\mathbb{E}[y] - \hat{y} \int p(y|x) dy = 0$$

$$\mathbb{E}[y] - \hat{y} = 0$$

$$\hat{y} = \mathbb{E}[y] = \boldsymbol{\phi}(x)^{\mathrm{T}} \mathbf{w}$$

Probabilistyczny model regresji liniowej z apriori

Model

Przyjmijmy następujący model:

$$p(y|x, \mathbf{w}, \sigma^2) = \mathcal{N}(y|\boldsymbol{\phi}(x)^{\mathrm{T}}\mathbf{w}, \sigma^2).$$

Przyjmijmy następujący rozkład a priori na parametry $\mathbf{w}$:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|0, \alpha^2\mathbb{I}) = \frac{1}{(2\pi)^{M/2}}\frac{1}{|\alpha^2\mathbb{I}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{w}^T(\alpha^2\mathbb{I})^{-1}\mathbf{w}\right\},$$

gdzie

$$det(\alpha^2\mathbb{I}) = \alpha^{2M}$$

$$(\alpha^2\mathbb{I})^{-1} = \frac{1}{\alpha^2}\mathbb{I}$$

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć rozkład a posteriori:

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć rozkład a posteriori:

$$p(\mathbf{w}|X,Y) = \frac{p(Y|X,\mathbf{w})p(\mathbf{w})}{p(Y|X)}$$

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć rozkład a posteriori:

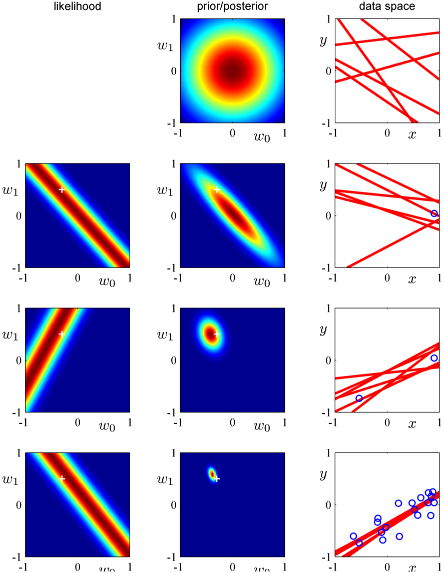
$$p(\mathbf{w}|X, Y) = \frac{p(Y|X, \mathbf{w})p(\mathbf{w})}{p(Y|X)}$$

po wstawieniu odpowiednich rozkładów:

$$p(\mathbf{w}|X, Y) = \frac{\mathcal{N}(\mathbf{w}|0, \alpha^2\mathbb{I}) \prod_{n=1}^{N} \mathcal{N}(y_n|\boldsymbol{\phi}(x_n)^{\mathrm{T}}\mathbf{w}, \sigma^2)}{p(Y|X)}$$

Probabilistyczny model regresji liniowej z apriori

Uczenie



Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć logarytm rozkładu a posteriori:

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć logarytm rozkładu a posteriori:

$$
\begin{aligned}
\ln p(\mathbf{w}|X,Y) &= \ln\left(\frac{1}{(2\pi)^{M/2}}\frac{1}{(\alpha^2)^{M/2}}\exp\left\{-\frac{1}{2\alpha^2}||\mathbf{w}||_2^2\right\}\right)\\
&\quad + \ln\left(\prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\boldsymbol{\phi}(x_n)^{\mathrm{T}}\mathbf{w})^2\right\}\right)\\
&\quad - \ln p(Y|X)\\
&= -\frac{M}{2}\ln 2\pi - \frac{M}{2}\ln\alpha^2 - \frac{1}{2\alpha^2}||\mathbf{w}||_2^2\\
&\quad - \frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}||Y-\Phi\mathbf{w}||_2^2\\
&\quad - \ln p(Y|X)
\end{aligned}
$$

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć gradient logarytmu rozkładu a posteriori względem parametrów $\mathbf{w}$:

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć gradient logarytmu rozkładu a posteriori względem parametrów $\mathbf{w}$:

$$\nabla_w \ln p(\mathbf{w}|X, Y) = -\frac{1}{2\alpha^2}2w - \frac{1}{2\sigma^2}(-2\Phi^\intercal Y + 2\Phi^\intercal\Phi\mathbf{w})$$

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć gradient logarytmu rozkładu a posteriori względem parametrów $\mathbf{w}$:

$$\nabla_w \ln p(\mathbf{w}|X,Y) = -\frac{1}{2\alpha^2}2w - \frac{1}{2\sigma^2}(-2\Phi^\intercal Y + 2\Phi^\intercal\Phi\mathbf{w})$$

i wyznaczyć optymalną wartość $\mathbf{w}$:

Probabilistyczny model regresji liniowej z apriori

Uczenie

Można wyznaczyć gradient logarytmu rozkładu a posteriori względem parametrów $\mathbf{w}$:

$$\nabla_w \ln p(\mathbf{w}|X, Y) = -\frac{1}{2\alpha^2} 2w - \frac{1}{2\sigma^2}(-2\Phi^\intercal Y + 2\Phi^\intercal \Phi \mathbf{w})$$

i wyznaczyć optymalną wartość $\mathbf{w}$:

$$-\frac{1}{2\alpha^2} 2\mathbf{w} - \frac{1}{2\sigma^2}(-2\Phi^\intercal Y + 2\Phi^\intercal \Phi \mathbf{w}) = 0$$

$$\Phi^\intercal Y - \Phi^\intercal \Phi \mathbf{w} - \frac{\sigma^2}{\alpha^2} \mathbf{w} = 0$$

$$\mathbf{w} = (\Phi^\intercal \Phi + \frac{\sigma^2}{\alpha^2} \mathbb{I})^{-1} \Phi^\intercal Y$$