Wykład 15

Regresja liniowa

Załóżmy, że dana jest macierz $X \in \mathbb{R}^{m\times 2}$ o postaci $X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$ oraz wektor $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$. Regresja liniowa polega na znalezieniu takiego wektora $\beta = [\beta_1, \beta_2]^T$ który minimalizuje sumę odległości punktów (x_i, y_i) od prostej $y = \beta_1 + \beta_2 x$. Innymi słowy: chcemy zminimalizować funkcję:

$$f(\beta_1, \beta_2) = \sum_{k=1}^{m} (\beta_1 + \beta_2 x_k - y_k)^2. \tag{1}$$

Obliczając pochodne cząstkowe funkcji f i przyrównując je do zera otrzymujemy układ równań określający zmienne β_1, β_2

$$\frac{\partial f}{\partial \beta_1} = 2 \cdot \sum_{k=1}^{m} ((\beta_1 + \beta_2 x_k - y_k) = 0, \quad \frac{\partial f}{\partial \beta_2} = 2 \cdot \sum_{k=1}^{m} ((\beta_1 + \beta_2 x_k - y_k) \cdot x_k = 0. \tag{2}$$

Pierwsze równanie daje wzór $\beta_1 = \sum_{k=1}^{m} y_k - \beta_2 \cdot \sum_{k=1}^{m} x_k$, który zwykle zapisujemy w zwięzłej postaci jako $\beta_1 = \bar{y} - \beta_2 \bar{x}$. Wstawienie tego wzoru do drugiego równania daje $0 = \beta_1 \bar{x} + \beta_2 \sum_{k=1}^{m} x_k^2 - \sum_{k=1}^{m} x_k y_k = \beta_1 \bar{x}\bar{y} - \sum_{k=1}^{m} x_k y_k + \beta_2 \left(\sum_{k=1}^{m} x_k^2 - m \cdot \bar{x}^2 \right)$. Stąd otrzymujemy rozwiązanie równań (2)

$$\beta_2 = \frac{\sum_{k=1}^{m} (x_k - \bar{x}) \cdot (y_k - \bar{y})}{\sum_{k=1}^{m} (x_k - \bar{x})^2}, \quad \beta_1 = \bar{y} - \beta_2 \bar{x}.$$

Zwróćmy uwagę, że wzór na β_2 ma prostą interpretację, mianowicie $\beta_2 = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}$.

Odległość punktu od prostej jest tutaj definiowana jako $|\beta_1 + \beta_2 x_k - y_k|^2$, nie jest to typowa odległość punktu od prostej, uzasadnienie (powody) nastąpi nieco później. Wzór (1) ma dodatkową interpretację: dla danych X, Y szukamy β taką, że $X \cdot \beta \approx Y$, przy czym β minimalizuje wartość wyrażenia $||X \cdot \beta - Y||_2^2$.

Niech zatem $g(\beta_1, \beta_2) = ||X \cdot \beta - Y||_2^2 = \beta^T X^T X \beta - 2 \cdot Y^T X \beta_1 + Y^T Y$. Bezpośrednim rachunkiem sprawdzamy, że

$$\begin{aligned}
&\beta X^T X \beta = [\beta_1, \beta_2] \begin{bmatrix} m & m\bar{x} \\ m\bar{x} & \sum x_k^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = m\beta_1^2 + 2m\bar{x}\beta_1\beta_2 + \left(\sum x_k^2\right) \beta_2^2, \\
&\beta^T X^T Y = Y^T X \beta = [\beta_1, \beta_2] \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = m\bar{y}\beta_1 + \left(\sum x_k y_k\right) \beta_2, \\
&Y^T Y = \sum y_k^2.
\end{aligned} \tag{3}$$

Równania (3) pozwalają na zapisanie funkcji $g(\beta_1, \beta_2) = ||X \cdot \beta - Y||_2^2$ w postaci skalarnej. Obliczając pochodne cząstkowe $\dfrac{\partial g}{\partial \beta_1}, \dfrac{\partial g}{\partial \beta_2}$ i przyrównując je do zera otrzymujemy równania

$$\begin{cases} 2m\beta_1 + 2m\bar{x}\beta_2 - 2m\bar{y} = 0, \\ 2\left(\sum x_k^2\right)\beta_2 + 2m\bar{x} - 2\sum x_k y_k = 0. \end{cases}$$

Rozwiązanie powyższego układu jest identyczne z rozwiązaniem układu (2) co dowodzi następujące

Twierdzenie 1. *Współczynniki równania regresji β minimalizują wartość wyrażenia $g(\beta_1, \beta_2) = ||X \cdot \beta - Y||_2^2 = \beta^T X^T X \beta - 2 \cdot Y^T X \beta_1 + Y^T Y$.*

Przejdźmy teraz do zadania aproksymacji średniokwadratowej. Dla danych punktów $x_1, \ldots, x_m$ określmy iloczyn skalarny funkcji $s(x), t(x)$ następująco: $\langle s, t\rangle = \sum_{k=1}^{m} s(x_k)\, t(x_k)$. Niech oprócz tego dana będzie funkcja $y(x)$, tzn. wektor $Y = [y_1, \ldots, y_m]^T$. Szukamy elementu najlepszej aproksymacji dla funkcji $y(x)$ z podprzestrzeni rozpiętej przez funkcje $\mathbb{1}$, $\mathbb{X}$.

Wiadomo[1], że element najlepszej aproksymacji jest określony przez układ równań normalnych

$$\begin{bmatrix} \langle \mathbb{1}, \mathbb{1}\rangle & \langle \mathbb{1}, \mathbb{X}\rangle \\ \langle \mathbb{X}, \mathbb{1}\rangle & \langle \mathbb{X}, \mathbb{X}\rangle \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \langle \mathbb{1}, \mathbb{Y}\rangle \\ \langle \mathbb{X}, \mathbb{Y}\rangle \end{bmatrix}. \tag{4}$$

Liczby α_1, α_2 to współczynniki elementu optymalnego względem funkcji $\mathbb{1}, \mathbb{X}$, tzn. $w_1(x) = \alpha_1 + \alpha_2 x$. Korzystając z postaci funkcji (wektorów) $\mathbb{1}$ oraz $\mathbb{X}$ układ równań (4) przybiera postać

$$\begin{bmatrix} m & m\bar{x} \\ m\bar{x} & \sum x_k^2 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} m\bar{y} \\ \sum x_k y_k \end{bmatrix}, \tag{5}$$

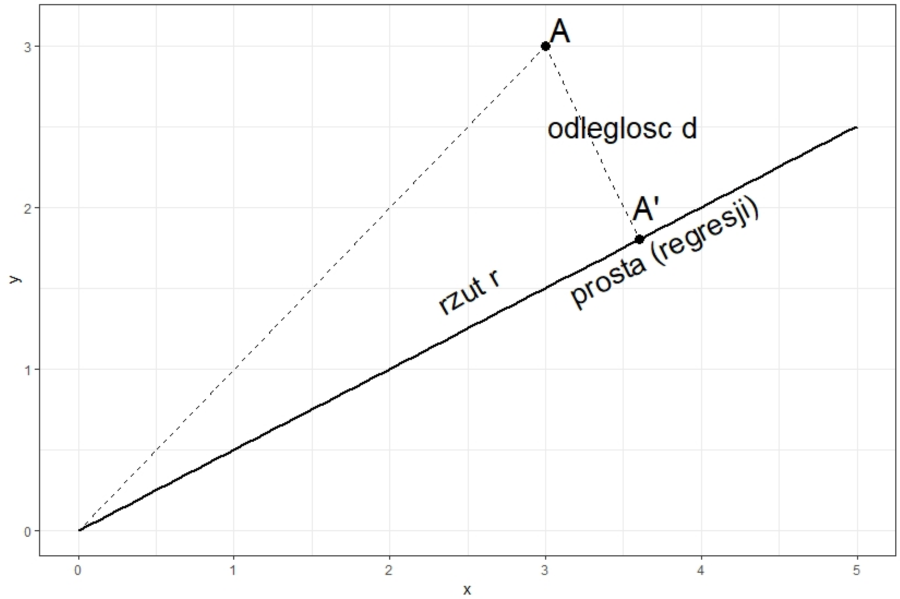
identyczną z układem równań określonym wzorami (2).

Przypomnijmy dodatkowo, że $w_n(x) \in \mathrm{lin}\{g_1(x), \ldots, g_n(x)\}$ jest elementem najlepszej aproksymacji dla funkcji $f(x)$ wtedy i tylko wtedy gdy $f - w_n \perp g_k$. Oznacza to, że:

1. prosta regresji zmiennej Y względem zmiennej X to pierwszy wielomian optymalny dla funkcji $y(x)$,
2. prosta regresji powstaje poprzez rzut prostokątny funkcji $y(x)$ na przestrzeń rozpiętą przez wektory (funkcje) $\mathbb{1}, \mathbb{X}$,
3. minimalizacja odległości pomiędzy punktami (x_i, y_i) a prostą regresji jest równoważna maksymalizacji długości rzutu punktów (x_i, y_i) na prostą (regresji).

[1] Analiza Numeryczna.

Krótkie objaśnienie do ostatniego z powyższych punktów. Rozważmy punkt A jak na rysunku poniżej. Prosta regresji zaznaczona jest linią ciągłą, natomiast linią przerywaną długość odcinka OA oraz odległość punktu A od prostej (odcinek AA'). Rzutujemy punkt w sposób prostokątny w związku z tym $OA^2 = r^2 + d^2$.



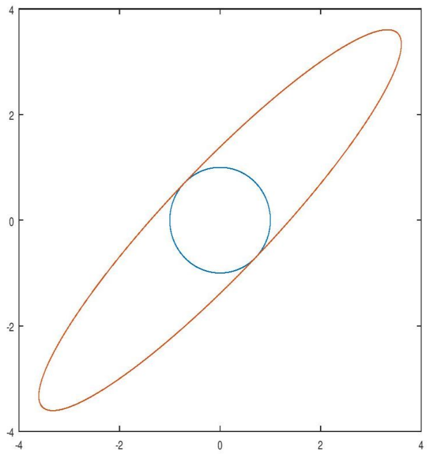
Ponieważ długość OA jest stała więc minimalizacja odległości d jest równoważna maksymalizacji długości rzutu r. Prosta oznacza w ogólności podprzestrzeń w której szukamy elementu najlepszej aproksymacji (dowolna przestrzeń zawiera wektor zerowy, stąd prosta przechodzi przez punkt 0). Punkt A (funkcję $y(x)$) rzutujemy prostokątnie na podprzestrzeń.

Przekształcenia liniowe

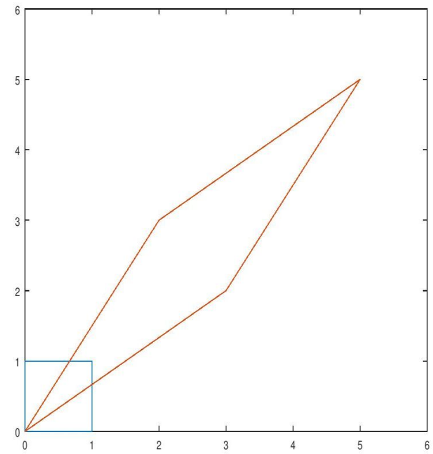
Przykład:

Rozpatrzmy macierz $A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$. Jej wartości własne to $\lambda_1 = 5$ oraz $\lambda_2 = 1$. Rysunek poniżej ilustruje obraz okręgu i kwadratu jednostkowego dla przekształcenia A. Obrazem okręgu jest elipsa o półosiach 5 oraz 1, obrazem kwadratu jest równoległobok o przekątnych $5\sqrt{2}, 1$. Tak elipsa jak i równoległobok są obrócone o 45^o.

Jest tak dlatego, że ortonormalne wektory własne macierzy A tworzą macierz obrotu o elementach $\frac{1}{2}\begin{pmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{pmatrix}$.



(a) Koło i jego obraz



(b) Kwadrat i jego obraz

Rozkład SVD – c.d.

Powracamy do rozkładu SVD.

Twierdzenie 2. *Niech $A \in \mathbb{R}^{m\times n}$. Wówczas istnieją macierze $U \in \mathbb{R}^{m\times m}$, $\Sigma \in \mathbb{R}^{m\times n}$ oraz $V \in \mathbb{R}^{n\times n}$ takie, że*

$$A = U\,\Sigma\,V^T,$$

gdzie macierze U, V są ortogonalne, macierz $\Sigma = diag(\sigma_1, \ldots \sigma_p)$, $\sigma_1 \geqslant \ldots, \sigma_p \geqslant 0$ oraz $p = min\{m, n\}$.

Przypomnijmy też, że macierze U, V to ortonormalne wektory własne dla AA^T oraz A^TA. Liczby $\sigma_1^2 \geqslant \ldots \geqslant \sigma_r^2 > 0$, gdzie $r \leqslant \min\{m, n\}$, to (niezerowe) wartości własne macierzy AA^T oraz A^TA.

Twierdzenie 3. *Niech u_k, v_k będą k-tymi kolumnami macierzy U, V z twierdzenia 2. Zachodzi wówczas równość $Av_k = \sigma_k u_k, \ \ k = 1, \ldots, p$.*

Dowód. Niech v_k będzie wektorem własnym odpowiadającym wartości własnej σ_k^2 macierzy A^TA. Jest zatem $A^TAv_k = \sigma_k^2 v_k$. Określmy wektor u_k następująco: $u_k = Av_k/\sigma_k$.

Sprawdzimy czy u_k jest wektorem własnym macierzy AA^T:

$$AA^Tu_k = \frac{1}{\sigma_k}A\left(A^TAv_k\right) = \frac{1}{\sigma_k}A\sigma_k^2 v_k = \sigma_k A v_k = \sigma_k^2 u_k.$$

□

Zadania

1. Niech $V_1 = A^T U_1 \bar{\Sigma}^{-1}$, gdzie $A \in \mathbb{R}^{m\times n}$, $U_1 \in \mathbb{R}^{m\times r}$, $\bar{\Sigma} = \operatorname{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}\right)$. Kolumny macierzy U_1 to ortonormalne wektory własne, odpowiadające niezerowym wartościom własnym macierzy AA^T. Sprawdzić, że $V_1^T V_1 = I_r$.
2. Niech $V_2 = A^T U_2 \bar{\Sigma}^{-1}$. Wykazać, że $V_2^T V_2 = I$ oraz $V_1^T V_2 = \mathbb{O}$.
3. Znaleźć macierz symetryczną 2×2 o wartościach własnych $\lambda_1 = 4, \lambda_2 = 0.25$. Czym jest obraz okręgu jednostkowego i kwadratu jednostkowego poprzez przekształcenie A?
4. Rozważmy macierz $A \in \mathbb{R}^{n\times n}$. Wykazać, że wartości własne macierzy A i A^T są takie same.
5. Niech $A \in \mathbb{R}^{m\times n}$. Co można powiedzieć o wartościach własnych macierzy AA^T i macierzy $A^T A$. Załóżmy dla przykładu, że $m \geqslant n$.
6. **2p.** Dana jest macierz $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}$.
   (a) Znaleźć wartości szczególne (σ_k) macierzy A.
   (b) Sprawdzić, że jeżeli $V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, to zachodzi $A^T A V = V \Sigma^T \Sigma$.
   (c) Korzystając z tw. 3 obliczyć wektory u_1, u_2.
   (d) Jak wybrać wektor u_3?