Regresjon og SVD

Oppgave 1

Anta at vi har sentrert datavektoren $\mathbf{x}$ til den sentrerte variabelen $C_n\mathbf{x}$ gitt ved

$$C_n\mathbf{x} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{x}$$

Vis at matrisen C_n er en projeksjon. Hva er gjennomsnittet til den sentrerte variabelen $C_n\mathbf{x}$?

Oppgave 2

Vi ser på et første ordens regresjonspolynom av to variable gitt ved

$$\mathbf{y} = \beta_{10}\mathbf{x}_1 + \beta_{01}\mathbf{x}_2 + \beta_0\mathbf{1}$$

Vis at dersom $\mathbf{x}_1$, $\mathbf{x}_2$ og $\mathbf{y}$ er sentrerte, så er $\beta_0 = 0$.

Oppgave 3

Vis at den empiriske kovariansmatrisen er positivt semidefinitt.

Oppgave 4

La

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

a) Finn den ortogonale diagonaliseringen av A^TA.

b) Finn den ortogonale diagonaliseringen av AA^T.

c) Finn singulærverdidekomposisjonen til A.

Her er kode for å finne SVD med `numpy` :)

Singulærverdidekomposisjon (SVD)

La $A \in \mathbb{R}^{m\times n}$ være en matrise. Da eksisterer ortogonale matriser $U \in \mathbb{R}^{m\times m}$ og $V \in \mathbb{R}^{n\times n}$, og en diagonalmatrise $\Sigma \in \mathbb{R}^{m\times n}$ slik at

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

Kolonnene i U er ortonormale egenvektorer til AA^T.
Kolonnene i V er ortonormale egenvektorer til A^TA.
Elementene langs diagonalen til Σ er *singulærverdier* $\sigma_i = \sqrt{\lambda_i}$ for $\lambda_i \neq 0$.

To viktige ting

A^TA og AA^T har samme ikke-null egenverdier. Kvadratroten av disse er singulærverdiene $\sigma_i = \sqrt{\lambda_i}$ som skal langs diagonalen til Σ. Sammenhengen mellom egenvektorer $\mathbf{v}_i$ til A^TA og egenvektorer $\mathbf{u}_i$ til AA^T er at $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$.
A^TA og AA^T er symmetriske matriser. Da vet vi at de er ortogonalt diagonaliserbare (altså de har en ortonormal egenbasis).

Hvordan finne SVD

1. Finn egenverdiene til A^TA. Kvadratroten av disse skal i synkende rekkefølge langs diagonalen til Σ.
2. Finn egenvektorene til A^TA. Normaliser disse. Dette blir kolonnene i V.
3. Finn egenvektorene til AA^T. Normaliser disse. Dette blir kolonnene i U.

Eksempel. La oss finne SVD til matrisen

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

A er en 2×3-matrise. Dermed er A^TA en 3×3-matrise og AA^T er en 2×2-matrise. AA^T og A^TA har samme ikke-null egenverdier, så vi velger å finne egenverdiene til AA^T, siden det er enklere å finne egenverdiene til en 2×2-matrise enn en 3×3-matrise.
Dermed beregner vi AA^T

$$AA^T = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Så finner vi egenverdiene til AA^T ved å løse $\det(AA^T - \lambda I) = 0$

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = (\lambda - 3)(\lambda - 1) = 0$$

Dermed får vi $\lambda_1 = 3$ og $\lambda_2 = 1$. Videre, finner vi egenvektorene til AA^T. Vi starter med $\lambda_1 = 3$

$$\left[\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right] \implies \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Så finner vi egenvektor til $\lambda_2 = 1$

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \implies \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Vi normaliserer disse (deler på lengden) og putter de inn som kolonner i U, slik at

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Videre, ønsker vi å finne egenvektorene til A^TA. Vi starter med å beregne A^TA

$$A^TA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Så finner vi egenvektoren til $\lambda_1 = 3$

$$\left[\begin{array}{ccc|c} -2 & 0 & 1 & 0 \\ 0 & -2 & 1 & 0 \\ 1 & 1 & -1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \implies \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

Så finner vi egenvektoren til $\lambda_2 = 1$

$$\left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \implies \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Siden A^TA er symmetrisk, vet vi at den har en ortogonal egenbasis for $\mathbb{R}^3$ ($\mathbb{R}^3$ siden $A^TA \in \mathbb{R}^{3\times3}$). Men vi har bare funnet to egenvektorer. Dermed må A^TA ha en egenverdi til, og denne *må* være 0, siden A^TA og

AA^T skal ha samme ikke-null egenverdier. Dermed finner vi egenvektoren til $\lambda_3 = 0$.

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \implies \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

Vi normaliserer $\mathbf{v}_1$, $\mathbf{v}_2$ og $\mathbf{v}_3$, og putter disse inn som kolonner i V, slik at

$$V = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}$$

Vi har to ikke-null egenverdier, nemlig 3 og 1. Dermed har vi to singulærverdier, nemlig $\sqrt{3}$ og 1, slik at matrisen Σ er gitt ved

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Dermed blir SVD

$$A = U\Sigma V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

Wolfram-alpha er enig:)