2 - 6 - HVA ER EN KATT? - LF

$\boxed{1}$ Jeg beregna faktisk den inverse i oppgave 2-9 i økt 2-4, det er bare å transponere A og så skalerer radene litt:

$$A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 1 & -\sqrt{2} & 1 \\ 1 & \sqrt{2} & 1 \end{pmatrix}$$

$\boxed{2}$ Her er det viktig å forstå at dersom P er en matrise med ortonormale kolonner, er

$$P^{-1} = P^T$$

og dette gjelder ikke for

$$V = \begin{pmatrix} 1 & 0 & 4 \\ 2 & 1 & -1 \\ 2 & -1 & -1 \end{pmatrix},$$

for

$$VV^T = V^TV = \begin{pmatrix} 1 & 2 & 2 \\ 0 & 1 & -1 \\ 4 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 4 \\ 2 & 1 & -1 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 18 \end{pmatrix},$$

men om vi normaliserer kolonnene og definerer

$$P = \begin{pmatrix} 1/3 & 0 & 4/\sqrt{18} \\ 2/3 & 1/\sqrt{2} & -1/\sqrt{18} \\ 2/3 & -1/\sqrt{2} & -1/\sqrt{18} \end{pmatrix},$$

er

$$PP^T = P^TP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Du kan nå sjekke selv at

$$A = P\Lambda P^T.$$

$\boxed{3}$ Vi beregner først det karakteristiske polynomet

$$\begin{aligned}
\det \begin{pmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{pmatrix} &= (2-\lambda)\left((2-\lambda)^2 - 1\right) - (2-\lambda-1) + (1-(2-\lambda)) \\
&= (2-\lambda)\left(3 - 4\lambda + \lambda^2\right) - (1-\lambda) - (1-\lambda) \\
&= (2-\lambda)(3-\lambda)(1-\lambda) - (1-\lambda) - (1-\lambda) \\
&= (1-\lambda)((2-\lambda)(3-\lambda) - 2) \\
&= (1-\lambda)(4 - 5\lambda + \lambda^2) \\
&= (1-\lambda)(1-\lambda)(4-\lambda)
\end{aligned}$$

og ser at vi har en enkel egenverdi $\lambda = 4$ og en dobbel egenverdi $\lambda = 1$. La oss begynne med den enkle egenverdien. Vi kunne gausset i vei, men siden vi elsker tall og observerer at tverrsummen av alle radene er 4, må egenvektoren være

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Så den doble. Siden

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

ser vi at alle egenvektorer ligger i planet til likningen

$$x_1 + x_2 + x_3 = 0$$

så for eksempel

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad \text{og} \qquad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

duger. Merk at begge er ortogonale på $\mathbf{v}_1$, men ikke på hverandre. Heldigvis er vi flinke i projeksjon, og ser at hvis vi bytter ut $\mathbf{v}_3$ med

$$\mathbf{v}_2 - \frac{\mathbf{v}_2^T \mathbf{v}_3}{\mathbf{v}_3^T \mathbf{v}_3} \mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

så er denne fortsatt egenvektor, men også ortogonal på både $\mathbf{v}_1$ og $\mathbf{v}_2$. Dersom vi normaliserer alle egenvektorer og setter dem opp som kolonner i en matrise

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

vil $A = VDV^T$, der

$$D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Merk at egenvektorene som korresponderte til forskjellige egenverdier automatisk kom ut ortogonale. Dette var ikke tilfeldig, for

$$\lambda_1 \mathbf{v}_1^T \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1^T \mathbf{v}_2$$

slik at

$$\lambda_1 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1^T \mathbf{v}_2$$

eller

$$(\lambda_1 - \lambda_2) \mathbf{v}_1^T \mathbf{v}_2 = 0.$$

Dersom A er symmetrisk og $\lambda_1 \neq \lambda_2$ må altså $\mathbf{v}_1^T \mathbf{v}_2 = 0$.

$\boxed{4}$ Dette er litt greit å vite, for det blir lettere å lese statistikkbøker om man er god i lineæralgebra.

$$\mathbf{1}^T\mathbf{1} = n$$

$$\mathbf{1}\mathbf{1}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

$$\mathbf{1}^T\mathbf{x} = n\overline{\mathbf{x}}$$

$\boxed{5}$ En projeksjon er en lineæroperator som tilfredsstiller $P = P^2$. Vi beregner

$$\begin{aligned} C_n^2 = & \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \\ = & I - \frac{2}{n}\mathbf{1}\mathbf{1}^T + \frac{1}{n}\mathbf{1}\mathbf{1}^T \\ = & I - \frac{1}{n}\mathbf{1}\mathbf{1}^T \end{aligned}$$

Gjennnomsnittet til den sentrerte variabelen er selvfølgelig null. Det er det som er poenget.

$\boxed{6}$ Med sentrerte variable der $\overline{\mathbf{x}} = \overline{\mathbf{y}} = 0$, får vi

$$\beta_1 = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x}} \qquad \text{og} \qquad \beta_0 = 0.$$

Nå kjenner vi igjen β_1 som projeksjonen av $\mathbf{y}$ på $\mathbf{x}$.

$\boxed{7}$ Vi har

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

altså

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \beta_{10}\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} + \beta_{01}\begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix} + \beta_0\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

La nå

$$\begin{pmatrix} x_{11} & x_{12} & 1 \\ x_{21} & x_{22} & 1 \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & 1 \end{pmatrix}.$$

Vi ganger med A^T fra venstre for å få normallikningene:

$$A^T\mathbf{y} = A^TA\beta$$

Skrevet ut på komponentform blir dette

$$\begin{pmatrix} \mathbf{y}^T\mathbf{x}_1 \\ \mathbf{y}^T\mathbf{x}_2 \\ n\overline{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}_1\|^2 & \mathbf{x}_1^T\mathbf{x}_2 & n\overline{\mathbf{x}_1} \\ \mathbf{x}_1^T\mathbf{x}_2 & \|\mathbf{x}_2\|^2 & n\overline{\mathbf{x}_2} \\ n\overline{\mathbf{x}_1} & n\overline{\mathbf{x}_2} & n \end{pmatrix}\begin{pmatrix} \beta_{10} \\ \beta_{01} \\ \beta_0 \end{pmatrix}$$

Dersom $\overline{\mathbf{x}_1} = \overline{\mathbf{x}_2} = \overline{\mathbf{y}} = 0$, blir dette mye enklere:

$$\begin{pmatrix} \mathbf{y}^T\mathbf{x}_1 \\ \mathbf{y}^T\mathbf{x}_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}_1\|^2 & \mathbf{x}_1^T\mathbf{x}_2 & 0 \\ \mathbf{x}_1^T\mathbf{x}_2 & \|\mathbf{x}_2\|^2 & 0 \\ 0 & 0 & n \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{01} \\ \beta_0 \end{pmatrix}$$

og vi ser lett at $\beta_0 = 0$, slik at likningssystemet blir

$$\begin{pmatrix} \mathbf{y}^T\mathbf{x}_1 \\ \mathbf{y}^T\mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}_1\|^2 & \mathbf{x}_1^T\mathbf{x}_2 \\ \mathbf{x}_1^T\mathbf{x}_2 & \|\mathbf{x}_2\|^2 \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{01} \end{pmatrix}$$

Et klassisk triks for å kjapt løse 2×2-matriser, er å kjenne til formelen

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

slik at

$$\begin{aligned} \beta &= \left(A^T A\right)^{-1} A\mathbf{y} \\ &= \frac{1}{\|\mathbf{x}_1\|^2 \|\mathbf{x}_2\|^2 - \left(\mathbf{x}_1^T\mathbf{x}_2\right)^2} \begin{pmatrix} \|\mathbf{x}_2\|^2 & -\mathbf{x}_1^T\mathbf{x}_2 \\ -\mathbf{x}_1^T\mathbf{x}_2 & \|\mathbf{x}_1\|^2 \end{pmatrix} \begin{pmatrix} \mathbf{y}^T\mathbf{x}_1 \\ \mathbf{y}^T\mathbf{x}_2 \end{pmatrix} \end{aligned}$$

I dataanalyse er det også vanlig å dele ut standardavvikene. Dette er det samme som å normalisere vektorer i lineæralgebra. Gjør vi det, blir det enda penere:

$$\beta = \frac{1}{1 - \left(\mathbf{x}_1^T\mathbf{x}_2\right)^2} \begin{pmatrix} 1 & -\mathbf{x}_1^T\mathbf{x}_2 \\ -\mathbf{x}_1^T\mathbf{x}_2 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{y}^T\mathbf{x}_1 \\ \mathbf{y}^T\mathbf{x}_2 \end{pmatrix}$$

$\boxed{8}$ Empirisk kovarians mellom to datasett $\mathbf{x}$ og $\mathbf{y}$ ser slik ut

$$\frac{1}{n-1} \sum_k (x_k - \overline{\mathbf{x}})(y_k - \overline{\mathbf{y}})$$

og dersom vi sentrerer, blir dette bare skalarproduktet mellom $\mathbf{x}$ og $\mathbf{y}$ delt på $n-1$. Hvis du ser nøye på $\frac{1}{n-1} X^T X$ vil du se at hvert element er den empiriske kovariansen mellom de forskjellige kolonnene i X siden $X^T X$ bare er skalarprodukter mellom kolonner i X.

$\boxed{9}$ Alle matriser på formen $A^T A$ er positivt definitte om de har full rang, og positivt semidefinitte om de ikke har det.[1] Dersom A har full rang, er

$$0 < \|A\mathbf{x}\|^2 = \mathbf{x}^T A^T A \mathbf{x}$$

så lenge $\mathbf{x} \neq \mathbf{0}$. Dersom A ikke har full rang, er

$$0 \leq \|A\mathbf{x}\|^2 = \mathbf{x}^T A^T A \mathbf{x}$$

siden det kan finnes $\mathbf{x} \neq \mathbf{0}$ slik at $A\mathbf{x} = \mathbf{0}$.

[1] Full rang betyr maksimalt antall lineært uavhengige rader eller kolonner: `https://en.wikipedia.org/wiki/Rank_(linear_algebra)`

$\boxed{10}$ PCA er en teknikk for å projisere dataskyen inn på en retning i rommet slik at variansen til det resulterende envariable datasettet blir maksimert:
`https://en.wikipedia.org/wiki/Principal_component_analysis`
Jeg har plaget dere med dette fordi det er sentralt i kunstig intelligens, og skal du skjønne noe av dét er du nødt til å forstå lineæralgebra og statistikk. La oss anta for enkelhets skyld at X kun har to kolonner. Vi ønsker å finne den retningen $\mathbf{p}$ i koordinatsystemet der datamengden er plotta slik at datamengden projisert på $\mathbf{p}$ har så stor varians som mulig. Produktet $X\mathbf{p}$ er en lang kolonnevektor der element k er projeksjonen av datapunkt k på $\mathbf{p}$, og dersom vi ønsker å maksimere summen av kvadratene av disse, må vi maksimere

$$\|X\mathbf{p}\| = \mathbf{p}^T X^T X\mathbf{p}.$$

Nå er det et standardresultat i numerisk analyse at om denne skal maksimeres, må man velge $\mathbf{p}$ som den egenvektoren til $X^T X$ med størst egenverdi.

$\boxed{11}$ Jeg er ikke ekspert på datatolkning, men får egenverdier på 36 og 28, så det betyr nok for alle praktiske formål ingen samvariasjon. Hvis det hadde vært stor forskjell på dem ville det betydd at det var mye variasjon når vi projiserte dataene på den ene egenvektoren og lite variasjon når vi projiserte på den andre.

$\boxed{13}$ De er ortogonalt diagonaliserbare siden de er reelle og symmetriske.

$\boxed{14}$ Dersom A har dimensjon $n \times p$, har $A^T A$ og V dimensjon $p \times p$ mens AA^T og U er $n \times n$.

$\boxed{15}$ Dersom $\mathbf{v}$ er en egenvektor til $A^T A$ med egenvektor $\lambda \neq 0$, kan vi gange likningen

$$A^T A\mathbf{v} = \lambda\mathbf{v}$$

med A og få

$$AA^T A\mathbf{v} = \lambda A\mathbf{v}$$

som sier at $A\mathbf{v}$ er en egenvektor til AA^T med den samme egenverdien. Dersom $\lambda = 0$ bryter dette sammen siden $A\mathbf{v} = 0$ og nullvektoren ikke klassifiserer som egenvektor. Resonnementet den andre veien er likt, og egenvektorene til AA^T kaller vi $\mathbf{u}$. Egenverdiene må være større enn null, siden

$$0 \leq \|A\mathbf{v}\|^2 = \mathbf{v}^T A^T A\mathbf{v} = \lambda\|v\|^2.$$

$\boxed{16}$ Beregningen i oppgaven over viser at dersom $\|\mathbf{v}\| = 1$ er

$$\|A\mathbf{v}\| = \sqrt{\lambda}.$$

$\boxed{17}$ I de foregående oppgavene har vi vist at

$$A\mathbf{v} = \sqrt{\lambda}\mathbf{u}$$

der $\mathbf{v}$ er en egenvektor til $A^T A$ og $\mathbf{u}$ er en egenvektor til AA^T, begge normaliserte og med egenverdi $\lambda \neq 0$. Setter vi opp alt dette i en matriselikning, får vi

$$AV = U\Sigma$$

der Σ er en diagonal og kvadratisk matrise med singulærverdiene på diagonalen. Dette kalles den **reduserte svd-faktoriseringen** til A. Dersom A^TA eller AA^T har egenverdier som er null kan vi utvide faktoriseringen til å inkludere de korresponderende egenvektorene, men disse er sjelden nyttige i praktiske anvendelser. Svd-faktoriseringen til

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

finner vi ved å beregne egenverdier og egenvektorer til

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

som er

$$\lambda_1 = 3 \qquad \mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \qquad \text{og} \qquad \lambda_2 = 1 \qquad \mathbf{v}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

og for

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

som er

$$\lambda_1 = 3 \quad \mathbf{u}_1 = \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix} \quad \text{og} \quad \lambda_2 = 1 \quad \mathbf{u}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \quad \text{og} \quad \lambda_3 = 0 \quad \mathbf{u}_3 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{pmatrix}$$

Du kan nå selv sjekke at

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

Dette kalles egentlig **redusert svd**. Vi kan inkludere den siste egenvektoren til AA^T, og skrive

$$A = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

som kalles **full svd**.

18 `https://folk.ntnu.no/mortano/python/katt/`