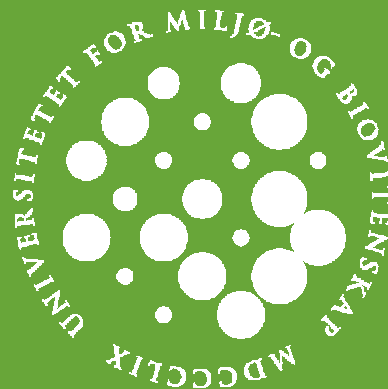


PPLS-DA: Egenskapsuttrekking med klassifikasjonsdata for bedret prediksjon og tolkbarhet

Kristian H. Liland og Ulf Indahl IKBM, UMB



Utgangspunkt

- "A twist to partial least squares" av Ulf Indahl om PPLS, <http://www3.interscience.wiley.com/cgi-bin/abstract/110557324/ABSTRACT>
- "Discrimination on latent components with respect to patterns. Application to multicollinear data" av Nocairi, Qannari m.fl. om PLS-DA <http://www.math.nccu.edu.tw/~jiangt/kuo/paper/paper/discriminant/9.pdf>

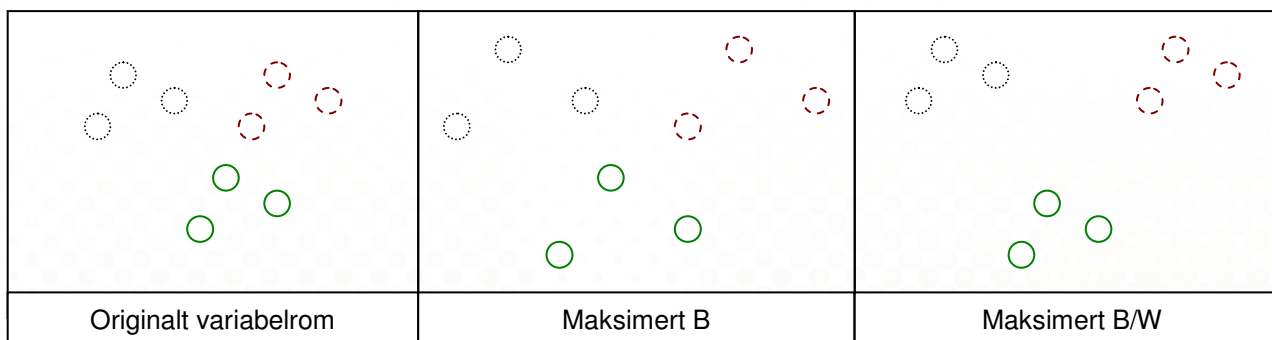
PPLS

- Ulf har gjennomgått PPLS
 - Korrelasjon og standardavvik potenseres
- Alf har gjennomgått utvidelse til multirespons
 - Bruker SVD på $X'Y$ for å hente ut ladningsvektor, \mathbf{w}
 - Faktoriserer ut standardavviket til X
 - Potenserer standardavviket og resten etter faktorisering
- PPLS-DA utvider disse med nye relasjonskriterier for å tilpasse den til klassifikasjon
- Noen av disse kriteriene er hentet fra PLS-DA

PLS-DA

- Bruker dummy-koding av responsvektoren:

$$\begin{bmatrix} \text{jente} \\ \text{gutt} \\ \text{jente} \\ \text{jente} \\ \text{gutt} \end{bmatrix} \Rightarrow Y = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 2 \end{bmatrix} \Leftrightarrow Y_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$
- Utnytter kovariansstrukturen til gruppene
 - Maksimering av mellom-gruppekovariansen, B
 - Maksimering av mellom-/innen-gruppekovariansen, $\frac{B}{W}$
- Den w som maksimerer $w'Bw$ kan vises å være den første egenvektoren til B som korresponderer til høyeste egenverdi
- Tilsvarende for $\frac{w'Bw}{w'Ww}$



- Tanken bak B og B/W ↑
- Foreløpig bare B som blir brukt til å lage ladningsvektorer, \mathbf{w}
 - Henter ut \mathbf{w} med SVD på B, definerer: $w_i = s_i \cdot |\text{rest}_i| \cdot \text{std}(\mathbf{x}_i)$
 - Faktoriserer variabelvis ut standardavviket til \mathbf{x}_i
 - Bytter ut korrelasjonen i relasjonskriteriet fra PPLS:

$$\mathbf{w}(\gamma) = K_\gamma [s_1 \cdot |\text{corr}(\mathbf{y}, \mathbf{x}_1)|^{\gamma/(1-\gamma)} \cdot \text{std}(\mathbf{x}_1)^{(1-\gamma)/\gamma}, \dots, s_p \cdot |\text{corr}(\mathbf{y}, \mathbf{x}_p)|^{\gamma/(1-\gamma)} \cdot \text{std}(\mathbf{x}_p)^{(1-\gamma)/\gamma}]$$
 ... med resten, "rest", etter å ha faktorisert ut standardavviket

$$\mathbf{w}(\gamma) = K_\gamma [s_1 \cdot |\text{rest}_1|^{\gamma/(1-\gamma)} \cdot \text{std}(\mathbf{x}_1)^{(1-\gamma)/\gamma}, \dots, s_p \cdot |\text{rest}_p|^{\gamma/(1-\gamma)} \cdot \text{std}(\mathbf{x}_p)^{(1-\gamma)/\gamma}]$$
- Både B og B/W blir brukt i bestemmelse av potenser
 - Mer om disse senere...

Full LDA/QDA

- For å skille mellom vanlig implementasjon av LDA/QDA og denne versjonen er denne kalt "full LDA/QDA"
- Forenkler ikke Bayes formel med naturlig logaritme for å kunne få ut aposteriori-sannsynlighetene for klassetilhørighetene til observasjonene
- Kan lage funksjoner av aposteriori-sannsynlighetene
 - Mer nøyaktige mål på suksessgraden til klassifisering enn bare å bruke feilraten eller suksessraten
 - Mulig å bruke som substitutt for relasjonsmål i potensering
- Tester tyder på at full QDA ofte bruker mellom 1/2 og 1/3 av tiden til vanlig QDA, mens vanlig LDA er marginalt raskere enn full LDA

Relasjonsmål for ladningsvektorer

- Vanligvis kovarians i PLS
- To hovedutgaver i PPLS-DA

– SVD på $X^T Y$ eller B . Faktoriserer ut standardavviket til X , som nevnt:

$$\mathbf{w}(\gamma) = K_\gamma [s_1 \cdot |\text{rest}_1|^{-\gamma/(1-\gamma)} \cdot \text{std}(\mathbf{x}_1)^{(1-\gamma)/\gamma}, \dots, s_p \cdot |\text{rest}_p|^{-\gamma/(1-\gamma)} \cdot \text{std}(\mathbf{x}_p)^{(1-\gamma)/\gamma}]$$

– Kandidater for \mathbf{w} fra hver av klassene, velger den som klassifiserer best, i følge

aposteriori-funksjonen:
$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^g (d_{i,j} - Yd_{i,j})$$

– Her ser vi på aposteriorivektorene som et forsøk på å gjenskape dummy utgaven av responsen

Relasjonsmål for potensering – Hvordan optimere valget av ladningsvektor $\mathbf{w}(\gamma)$

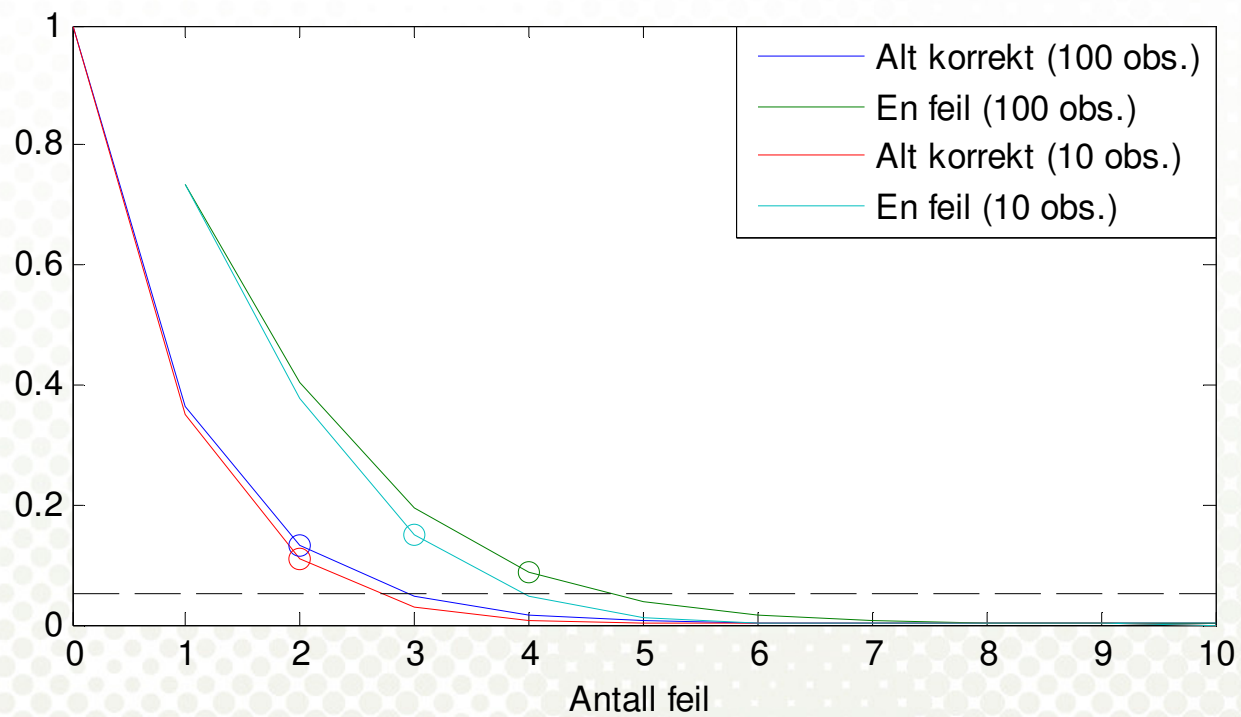
- Mange muligheter
- Oppgaven nevner:
 - Produkt av korrelasjon som i PPLS2
 - Maksimering av B ved $\mathbf{w}(\gamma)'B\mathbf{w}(\gamma)$
 - Maksimering av B/W ved $(\mathbf{w}(\gamma)'B\mathbf{w}(\gamma))/(\mathbf{w}(\gamma)'W\mathbf{w}(\gamma))$
 - Optimering med LDA og aposteriorifunksjon
Utfører en klassifikasjon med full LDA og minimerer $\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^g (d_{i,j} - Yd_{i,j})$
 - De to siste kandidatene er de mest interessante

Modellens godhet

- Den tradisjonelle måten å bestemme hvor god en klassifikasjonsmodell er er å se på andelen korrekt klassifiserte eller feilklassifiserte
- Et mål som ligner mye på feilandelen får en ved å plukke ut aposterioriverdiene som korresponderer med den egentlige klassen til hver observasjon og tar gjennomsnittet av disse
- Aposteriorifunksjonen $\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^g (d_{i,j} - Y_{d_{i,j}})$ har den fordel at den bruker flere komponenter på å nå 0 enn feilandelen

Enklere modell

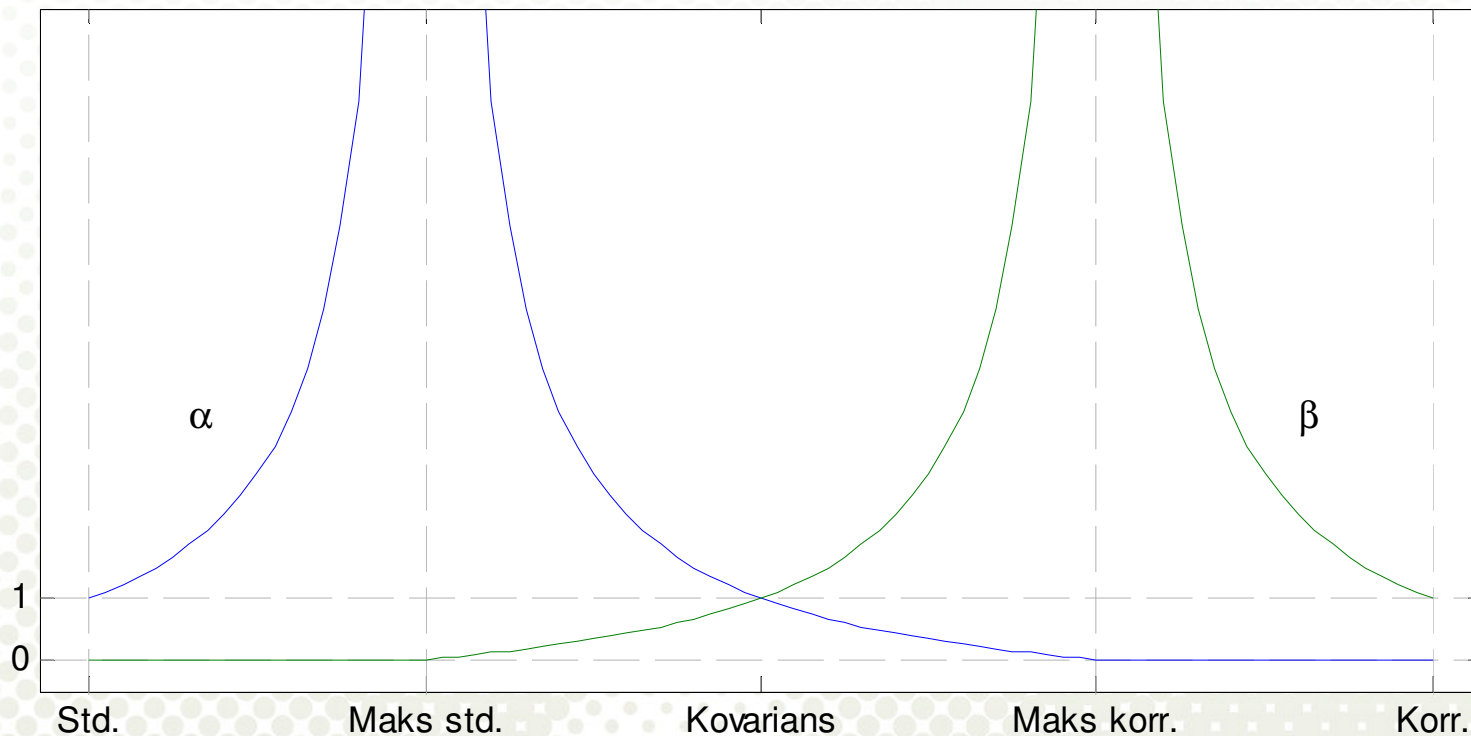
- I regresjon kan en bruke χ^2 -fordelingen til å finne den enkleste modellen som ikke er signifikant dårligere enn den beste
 - bruker kvadratavviket mellom modellens resultat og observasjonene
- I klassifikasjon, der observasjoner enten er korrekt eller galt klassifisert, er binomisk fordeling mer naturlig, hvis en bruker feilandel som mål på godhet
- Vi bruker kumulativ binomisk fordeling der feilandelene fra de enklere kandidatmodellene spiller rollen som p hver sin gang. Finner sannsynligheten for å få antallet feil fra den beste modellen
- Velger et signifikansnivå α som grense for hvilke modeller som skal godtas
- Sørger for en mer robust modell og reduserer faren for overtilpasning



- Strengere med færre observasjoner
- Mindre strengt med flere feil i enkleste modell

Potensområde

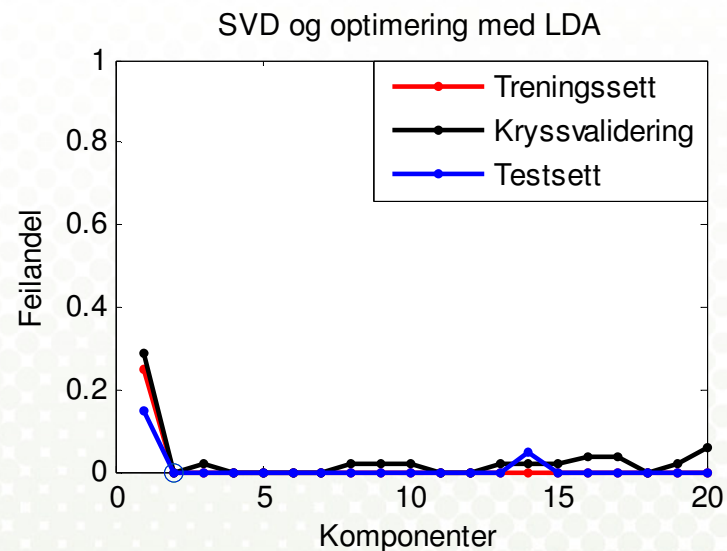
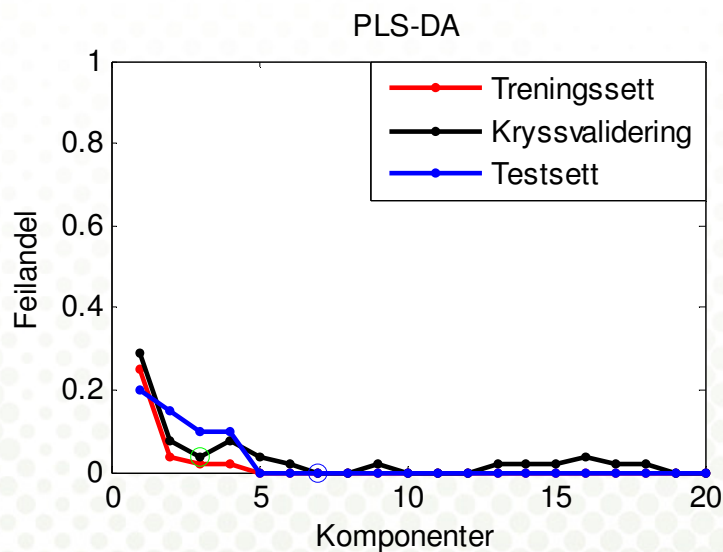
- I Noen tilfeller lønner det seg å standardisere data før de blir utsatt for en egenskapsuttrekkingsmetode. Det er ikke alltid så lett å vite om dette vil lønne seg på forhånd
- Ved å utvide potensområdet fra PPLS tar vi høyde for dette



NIR-data

- Et forsøk utført ved UMB
- 6 forskjellige typer oljer
- 72 observasjoner med nærinfrarød spektroskopi
- 351 bølgelengder fra 1100 til 2500 nanometer
- Delt i treningssett på 52 observasjoner og testsett på 20 observasjoner

Komponenter og feilandeler

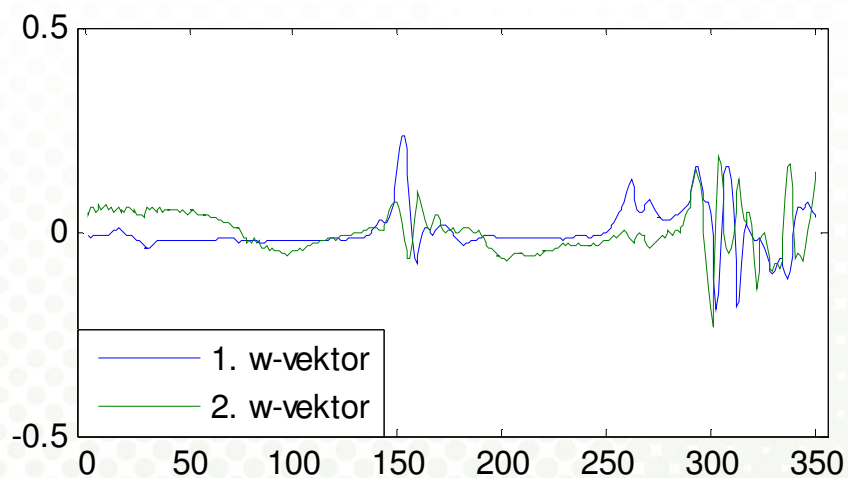


	Treningssett		Kryssvalidering				Testsett		
	pc	feil	pc	feil	pc	feil	pc	feil	
PLS-DA	5	0	7	0	3	2	5	0	
Beste klasse	B/W	2	0	4	0	2	2	2	0
	Opt. LDA	3	0	8	0	4	1	3	0
SVD	B/W	4	0	4	0	2	2	4	0
	Opt. LDA	2	0	2	0	2	0	2	0
Vektet SVD	B/W	3	0	3	0	3	0	4	0
	Opt. LDA	3	0	3	0	3	0	3	0
		pc	feil	pc	feil	pc	feil	pc	feil

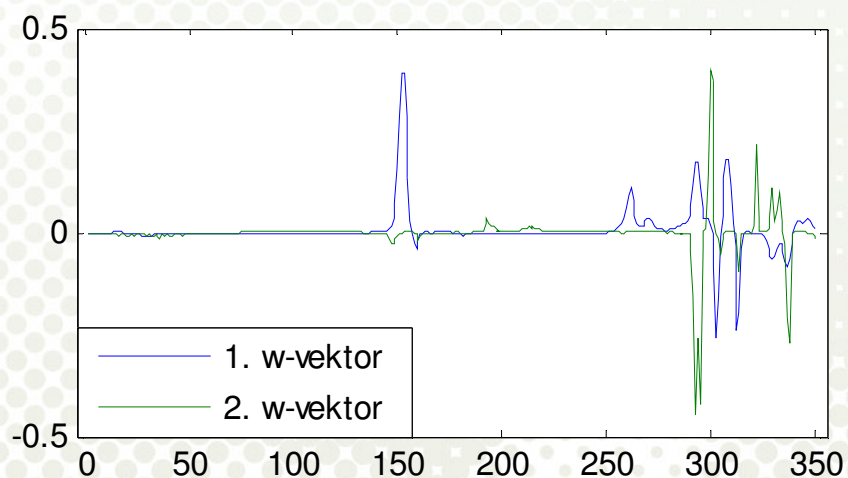
$\gamma=[0,1]$



Ladningsvektorer



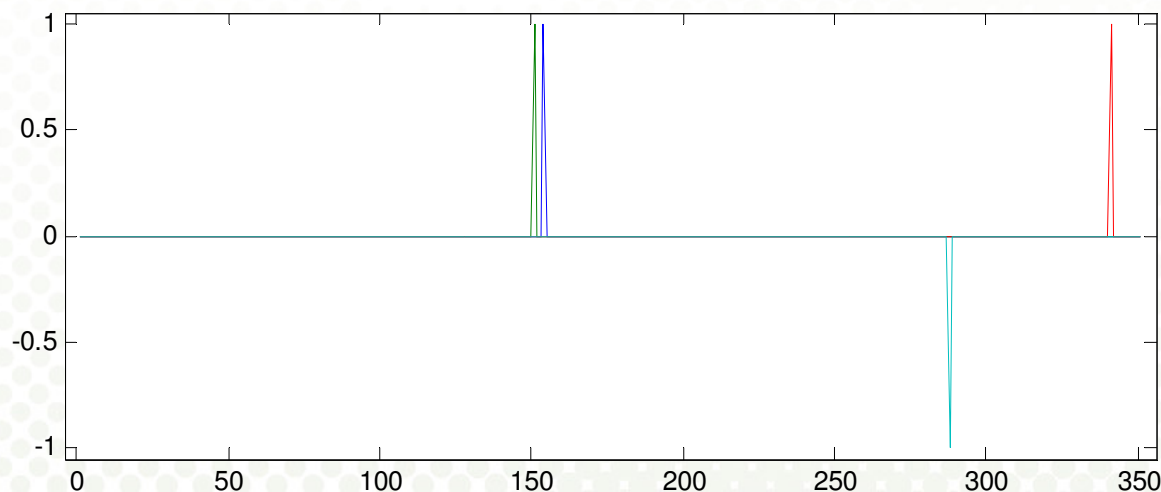
● PLS-DA



● SVD på $X'Y$

● Optimering med LDA

Fokus på enkeltvariabler



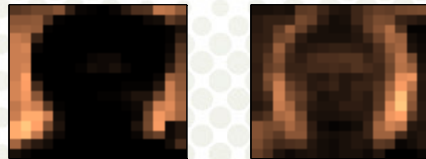
- PPLS-DA med SVD på X'Y med variabelseleksjon for korrelasjon
- Etter 4 komponenter er alt korrekt klassifisert med treningssett, testsett og kryssvalidering
 - Variabel 151, 154, 288 og 341
 - Tilsvarende 1704, 1716, 2252 og 2464 nm

Ansiktsgjenkjenning

- 30 personer
- 10 gråtonebilder av hver person med litt forskjellig mimikk og annen variasjon mellom bildene
- - 2 ødelagte bilder = 298 observasjoner
- Bilder à 16x16 piksler => 256 variable



- Ladningsvektorer kan visualiseres som bilder



Komponenter

	Trenings- sett		Kryssvalidering				
			Beste		Enkleste		
PLS-DA	9	0	9	1	7	3	
Beste klasse	B/W	5	0	16	1	11	4
	Opt. LDA	6	0	20	0	12	1
SVD	B/W	9	0	9	0	7	1
	Opt. LDA	9	0	9	0	8	1
Vektet SVD	B/W	9	0	9	0	7	2
	Opt. LDA	9	0	13	1	8	3
		pc	feil	pc	feil	pc	feil

$\gamma=[0,1]$

	Trenings- sett		Kryssvalidering				
			Beste		Enkleste		
PLS-DA	9	0	9	1	7	3	
Beste klasse	B/W	6	0	9	0	9	0
	Opt. LDA	7	0	16	2	9	6
SVD	B/W	9	0	9	0	7	1
	Opt. LDA	9	0	9	0	7	2
Vektet SVD	B/W	9	0	9	0	7	2
	Opt. LDA	9	0	9	1	8	3
		pc	feil	pc	feil	pc	feil

$\gamma=[0.1,0.9]$