

# ITMAL Øvelser – Probabilistiske modeller

## Øvelse 1 :

I skal modellere "height-weight" data fra den tidligere øvelse på kvinder og mænd med probabilistisk modeller.

Fit en multivariat (2D) normalfordeling til hhv. kvinder og mænds data – dvs. for hver klasse estimer middelværdi vektor  $\mu$  og covarians matrix  $\Sigma$ . I kan benyt metoden Quadratic Discriminant Analysis (QDA) til at estimere såkaldt "class conditional" sandsynlighed  $P(X|y=k)$  (sandsynlighedstæthed i punkt  $X=[\text{højde, vægt}]$  for hver klasse  $k$  (mænd og kvinder)) – metoden gemmer også middelværdi vektor  $\mu$  og covarians matrix  $\Sigma$  i QDA-objektet for hver klasse.

- Plot surface/contour af de to klassers pdf (probability distribution function).
- Hvad er pdf værdien i "centrum" (ved middelværdien) for mænd ?
- Argumentér for hvorfor en højde på 190 cm og vægt på 30 kg er en outlier udfra pdf-værdier (sund fornuft må ikke bruges.. metoden kan tænkes anvendt på features som ikke er "forståelige" såsom pixel-værdier i et billede). Metoden kan altså benyttes til outlier detection.

I skal nu benytte  $P(y=k|X)$  – dvs. sandsynlighed for klasse  $k$  betinget på en sample  $X$ , dvs. når vi er givet  $[\text{højde, vægt}]$ . Det er jo hvad vi ønsker at bruge i en klassifiser – dette er den primære brug af LDA og QDA.

- Plot decision boundaries (udgangspunkt i eksempel nedenfor) for LDA og QDA. Hvad er træningsfejlen ?
- Hvad er sandsynligheden  $P(y=\text{kvinde} | X=[170\text{cm}, 80\text{kg}])$  og  $P(y=\text{mand} | X=[170\text{cm}, 80\text{kg}])$  ?
- Optional : Vi kan ønske at være mere sikre i vores valg – dvs. såkaldt "reject option", hvor vi kun ønsker at benytte mere sikre samples. Hvad er træningsfejlen hvis vi kun klassificerer i klasse  $k$  når  $P(y=k|X) > 0,7$  (istf. 0,5 som normalt)

Tips :

Linear Discriminant Analysis (LDA) og Quadratic Discriminant Analysis (QDA) intro - [https://scikit-learn.org/stable/modules/lda\\_qda.html#lda-qda](https://scikit-learn.org/stable/modules/lda_qda.html#lda-qda)

QDA - [https://scikit-learn.org/stable/modules/generated/sklearn.discriminant\\_analysis.QuadraticDiscriminantAnalysis.html#sklearn.discriminant\\_analysis.QuadraticDiscriminantAnalysis](https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.QuadraticDiscriminantAnalysis.html#sklearn.discriminant_analysis.QuadraticDiscriminantAnalysis)

Surface plots - <https://problemsolvingwithpython.com/06-Plotting-with-Matplotlib/06.16-3D-Surface-Plots/>

## **Øvelse 2 :**

Afprøv LDA og QDA på jeres egne data.