# Oplæsning

## Formelsamling

### Sandsynlighed

#### Bayes theorem

posterior distribution  
 likelihood   
 prior distribution  
 normaliserings konstant

Indholdsfortegnelse

[Eksamens udtag 1](#_Toc187252645)

[Formelsamling 1](#_Toc187252646)

[Sandsynlighed 1](#_Toc187252647)

[Opgave 1. 2](#_Toc187252648)

[Opgave 2. 4](#_Toc187252649)

[Opgave 3. Klassifikation 7](#_Toc187252650)

[Opgave 4. Principal Component Analysis ( PCA - Forelæsning 8 ). 10](#_Toc187252651)

# Eksamens udtag

## Opgave 1.

Measurements of a 1-dimensional continuous variable belonging to two classes (blue and red) follow the probability distributions illustrated in Figure 1. The class probability distributions are equal to

Et billede, der indeholder diagram, linje/række, Kurve, Rektangel

Automatisk genereret beskrivelse

Class distributions of measurements of a 1-dimensional variable x. Blue corresponds to and red corresponds to

The values in the figure are equal to:

1. What is the probability of observing a measurement with value of

Her skal jeg tage in mente, at sandsynlighederne for klassifikationerne ikke er den samme. Derfor skal den enkelte’s sandsynlighed for optræden ganges på de summer af sandsynlighed, som jeg kommer til at finde.   
Summen af sandsynlighederne, kan jeg så bruge til at beregne hvor sandsynligt noget er for at ske.   
   
Men jeg kan også vende den om og sige, hvad er sandsynligheden for noget over 9.   
   
============================  
   
============================

1. What is the probability of a measurement belonging to to have a value of

So it’s a conditional probability. We then don’t care about C1’s effect on all the measurements.

Seen from within the C1

1. What is the probability of a measurement with a value of

Den har jeg allerede regnet.

1. Et billede, der indeholder diagram, linje/række, Kurve, Plan

   Automatisk genereret beskrivelseDescribe a way to classify a measurement with value of to one of the two classes using the probability values provided in this exercise



Hvis

Jeg kan vende betingelserne om vha. Bayes theorem.



Som I stedet for at sige, givet den her klassifikation, hvad vil procenten være for den her x, så siger jeg så:   
Givet den her x større eller mindre end a, hvad er sandsynligheden for, at det er klasse 1 eller 2?   
Så kan jeg stille spørgsmålet til at være:   
   
Jeg har:

Og jeg har at

Så kan jeg tage maks til det, delt med afstanden.

Så er det klart, at det nye element skal tilhøre klassifikation 1.

## Opgave 2.

Provide answers to the following questions:

1. What is conjugacy, where do we use it in probability-based classification, and why do we use it? (Score: 3 x 10%)

Conjugacy is when to probabilities from the same distribution is combined in such a way, the resultant probability is of same distribution.  
Using likelihood we can make models to have the best fit to our datapoints. But when we make the model to exact, overfitting becomes an issue.   
To overcome this, we take the likelihood and some prior distribution to get our next estimate of a model. This is where conjugacy is used, as we need a prior distribution and the likelihood to be of the same type of distribution, to make the posterior distribution be of same type of distribution.

1. When do we say that a dataset is formed by independent and identically distributed data? (Score: 10%) ( i.i.d )

We do that, when the datapoints aren’t dependent of each other and the points are from same distribution. That comes with some properties, like being able to distributing the points. For the probability of datapoints in a set, it can be described as

1. A set of 2-dimensional data points follow the distribution in Eq (1):

The parameters and of the distribution are equal to:

Answer the following questions:

**(Score: 20%)** Given three data points: , add the (in)equality signs in the following and describe why these signs are used (provide a qualitative explanation and do not calculate the values of p(x) for the three points):

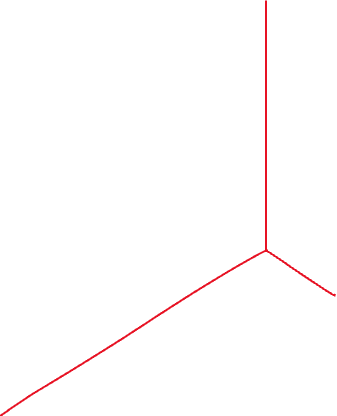


, therefore the first dimension is affected by c and the second dimension is affected by f.

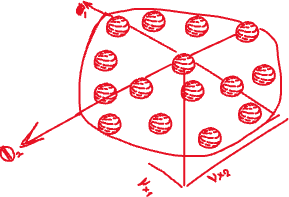
**(Score: 20%)** We decide to transform the data points to data points by applying data whitening.

Et billede, der indeholder tekst, Font/skrifttype, kvittering, hvid

Automatisk genereret beskrivelseFor at det skal ske, så skal jeg lave en transformation på datasættet, så at   
   
Og at   
   
Og til det ønsker jeg at dele min covariance matrice op i:   
Hvor grafisk kan repræsentere de retninger som min data spreder sig i. er så min egenværdier, som fortæller hvor meget   
mit data spreder sig på de akser. *( Det her bruges ofte til at fjerne   
afhængighed i covariancen, altså værdierne på den modsatte  
diagonal )*.



Løsning for   
   
Så lambda er bare min covariance matrice.



Normal vil jeg løse for for at finde vektorerne, men da det er en diagonal matrice, så er disse elementer bare 1 til den respectative værdi. Dette er for at

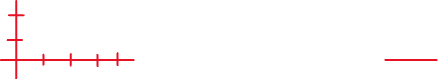
Og da   
Og at jeg ønsker en covariance som er identitets matricen:   
   
, da min er en diagonal basis matrice.

## Opgave 3. Klassifikation, K-nearest neighbour, K-Nearest centroids & bayes classifier.

A classification problem is formed by two classes. We are given a set of 2-dimensional data

Each belonging to one of the two classes, as indicated in the class label vector

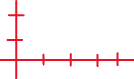
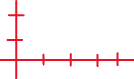
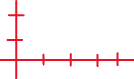
Using the above (training) vector and the corresponding class labels, classify the following vectors:



Questions:

1. (Score: 15%) The Nearest Class Centroid classifier (NCC).

Undervisning 9 bliver det her nævnt.



Så punkternes afstand til centrummerne bestemmer hvilken klassifikation den tilhører. Den mindste afstand vinder.   
Grafisk kan jeg fra start se at

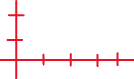
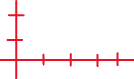
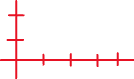
Men den sidste må jeg beregne.

Og det er derfor den er svær at klassificere ud fra sketch. Afstanden er næsten lige med hinanden.   
Men jeg kan altså konkludere at afstanden til anden klassifikation er mindst, derfor   
   
=============  
   
=============

1. (Score: 15%) The Nearest Neighbor classifier using only one neighbor classifier (1-NN)

For hvert punkt:

* Kig på de tættest K antal naboer, for hver klassifikation. Den klassifikation hvor summerne af distancen er mindst, er den klasse som den bliver klassificeret til at være.



Uklassificerbar, samme afstand til begge naboer.

=============  
   
=============

1. (Score: 30%) The Bayes-based classification scheme, where:

Ahh, jeg havde lige brug for at se hvad denne specifikke betegnelse står for. Jeg troede først, at det kun var normal fordeling, men det er bare den generelle betegnelse ud fra bayes theorem.

Og   
 er posterior distribution.   
 er likelihood.  
 er prior distribution.   
 bruges som en normaliserings konstant.

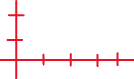
Og jeg kan godt huske fra undervisning, at vi kan dele det sådan op. Og vi har jo beregnet for de her flere gange.   
Her tror jeg kun, at vi ønsker likelihood, og er ligeglade med posterior og prior.   
Det sidste jeg har fået givet er:

Where  is the class mean vector of class  and:

Og min beslutning træffer jeg ud fra,

Så derfor får jeg at:

Så jeg får at   
=============  
   
=============



1. (Score: 40%) Compare (qualitatively) the decision functions obtained by using the NCC classifier and the above Bayes-based classifier.

De er nemlig tæt på hinanden. Jeg fik den samme klassificering ud af det, og er det klart?   
Afstands formlen er målt på den samme måde. Og det er egentlig kun nævneren der er anderledes. Nævneren er den samlede afstand til alle klasser, her de to klasser.   
Så nævneren sørger egentlig bare for at få normalizeret distancerne, så jeg med bayes based classifier beskriver sandsynligheden af, at punktet tilhører en given klasse. For NCC tager man bare den største værdi.   
Bayes classifier er dog skridtet tættere på at undgå overfitting, når man går videre til .

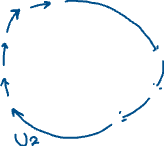
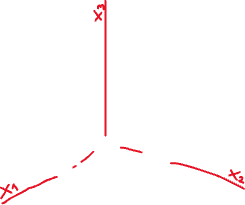
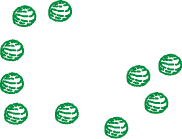
## Opgave 4. Principal Component Analysis ( PCA - Forelæsning 8 ).

Consider that a dataset is formed by *N* vectors . Let us assume that the dataset is zero-centered, and that we apply Principal Component Analysis to determine the first principal component using the maximum variance criterion:

where  is the total scatter matrix of the data . Show that the first principal component calculated by optimizing  is equivalent to the solution of the minimum reconstruction error defined as follows:

where

Idéen er, at vi skal finde en retning med *M* dimensioner som er mindre end   
*D* dimensioner som datapunkterne tager. Datapunkterne tager så den   
nye retningsfunktion samt en viden om fejlen som findes. Målet er så at gøre  
det uden at miste alt for meget præcision.



Til at lykkedes med det, så bliver der først maximeret for variansen i .   
Når man så skal estimere punkterne ud fra den ny reducerede dimension,  
så minimere man for fejlen i .

Så hvad skal jeg gøre?   
Okay jeg havde aldrig regnet den her ud selv.   
Det er linear algebra gymnastik.

Der bliver brugt at:   
   
Trace of Matrice hedder metoden og den tager summen af elementerne på diagonalen.

Egenskaberne:

Og at   
Så

Så forstår jeg ikke hans næste steps, men har gør:   
Et billede, der indeholder tekst, Font/skrifttype, kvittering, hvid

Automatisk genereret beskrivelse