# Statistical Learning and Machine Learning

Indholdsfortegnelse

[Statistical Learning and Machine Learning 1](#_Toc187347826)

[Formelsamling 2](#_Toc187347827)

[Sandsynlighed 2](#_Toc187347828)

[Bayes theorem 2](#_Toc187347829)

[Beviser 3](#_Toc187347830)

[Udledning af koefficienter for at til curve fitting ud fra error square funktionen. 3](#_Toc187347831)

[Opgaver 5](#_Toc187347832)

[Eksamens opgave 5](#_Toc187347833)

[Opgave 1. Sandsynlighed ud fra histogram. 5](#_Toc187347834)

[Opgave 2. 7](#_Toc187347835)

[Opgave 3. Klassifikation, K-nearest neighbour, K-Nearest centroids & bayes classifier. 10](#_Toc187347836)

[Opgave 4. Principal Component Analysis ( PCA - Forelæsning 8 ). 13](#_Toc187347837)

[Eksamensopgave 2024 15](#_Toc187347838)

[Tema 1 - Regression. 15](#_Toc187347839)

[Tema 2 - Klassifikation 17](#_Toc187347840)

[Tema 3 - Neural netværk. 20](#_Toc187347841)

[Opgave 1.1 Nyt forsøg. Udledning af w koefficienter ud fra error sum of squares. 23](#_Toc187347842)

[Opgave 1.1. Gammel forsøg 25](#_Toc187347843)

[Opgave 1.3 Kasser i forskellige farver, hver med forskelligt antal af samme frugter 27](#_Toc187347844)

[Opgave 1.3 for ny 29](#_Toc187347845)

[Opgave 1.7 - Den gaussiske distribuering og integral bevis. 30](#_Toc187347846)

[Opgave 1.10 31](#_Toc187347847)

[Opgave 1.12 32](#_Toc187347848)

[Opgave 6.5. Rigtig kernel eller ej? ( Linearitets princip og to ulineære funktioner til ny ulineær funktion ) 33](#_Toc187347849)

[Opgave 6.6. Rigtig kernel eller ej? Polynomial og eksponentielle transformationer. 33](#_Toc187347850)

[Opgave 6.7. Rigtig kernel eller ej? Polynomial og eksponentielle transformationer. 35](#_Toc187347851)

[Opgave 6.8-9 Rigtig kernel eller ej? Har ikke fortsat 35](#_Toc187347852)

[Opgave kapitel 6. Kernel PCA 36](#_Toc187347853)

[Opgave kapitel 6. Nolinear classification 37](#_Toc187347854)

[Opgave 7.2. Show that, if the 1 on the right hand side of the constraint (7.5) is replaced by some arbitrary constant , the solution for the maximum margin hyperplane is unchanged. 39](#_Toc187347855)

## Formelsamling

### Sandsynlighed

#### Bayes theorem

posterior distribution  
 likelihood   
 prior distribution  
 normaliserings konstant

## Beviser

### Udledning af koefficienter for at til curve fitting ud fra error square funktionen.

*Lavet i opgave 1.1*

Jeg skal få afledt den, sat den lige med nul for så at finde koefficienterne.   
Jeg afleder i forhold til koefficienterne ,og sætter det til 0, fordi hvorfor ikke?!

Den yderste summering varierer ikke i forhold til .

Der findes en sammenhængen funktion i den variende del.

afhænger af , så så længe at kan jeg ikke sige, at det afledte vil være anderledes fra 0. I det tilfælde at , så gælder der at:

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

Automatisk genereret beskrivelse=============================

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

Hvilket var hvad jeg skulle bevise.

## Opgaver

### Eksamens opgave

#### Opgave 1. Sandsynlighed ud fra histogram.

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:

###### 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.   
   
============================  
   
============================

###### 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

###### What is the probability of a measurement with a value of

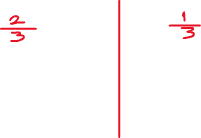
Den har jeg allerede regnet.

###### Describe a way to classify a measurement with value of to one of the two classes using the probability values provided in this exercise

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

Automatisk genereret beskrivelseHvis

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:

###### 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.

###### 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

###### 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):

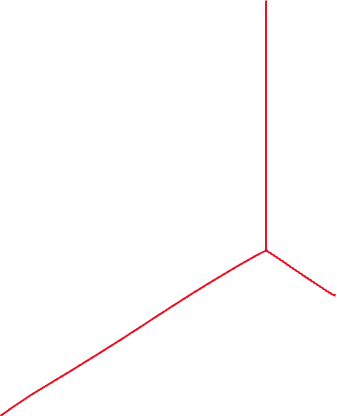


, therefore the first dimension is affected by c and the second dimension is affected by f.

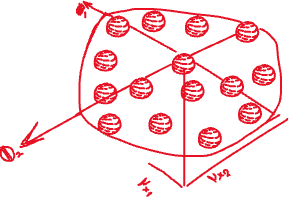
**(Score: 20%)** We decide to transform the data points to data points by applying data whitening.

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

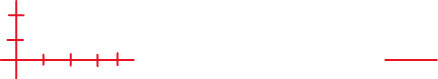
Og da   
Og at jeg ønsker en covariance som er identitets matricen:   
   
, da min er en diagonal basis matrice.

#### 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

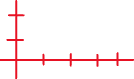
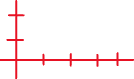
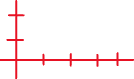
Using the above (training) vector and the corresponding class labels, classify the following vectors:



Questions:

###### (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

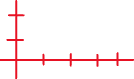
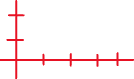
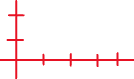
Men den sidste må jeg beregne.

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

###### (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.

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

###### (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.

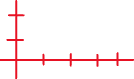
Og jeg kan godt huske fra undervisning, at vi kan dele det sådan op. Og vi har jo beregnet for de her flere gange.   
Her tror jeg kun, at vi ønsker likelihood, og er ligeglade med posterior og prior.   
Det sidste jeg har fået givet er:

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



###### (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 .

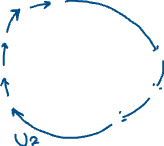
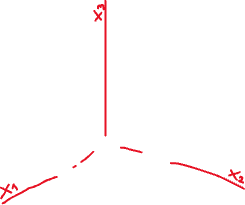
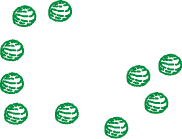
#### Opgave 4. Principal Component Analysis ( PCA - Forelæsning 8 ).

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

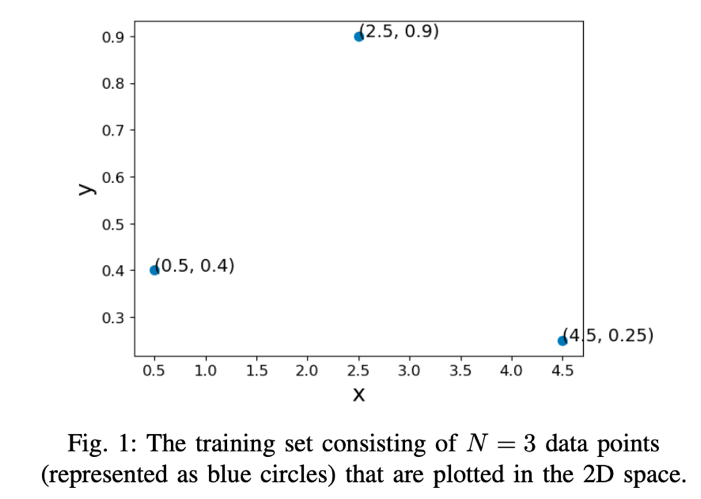
Så forstår jeg ikke hans næste steps, men har gør:   
Et billede, der indeholder tekst, Font/skrifttype, kvittering, hvid

Automatisk genereret beskrivelse

### Eksamensopgave 2024

#### Tema 1 - Regression.

Et billede, der indeholder tekst, skærmbillede, diagram, linje/række

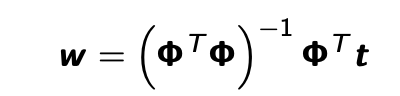
Automatisk genereret beskrivelse

###### Min basis functioner , sæt så design matrice up, og brug den til at beskrive the normal equations for least squares. ( 30 % )

Linear regression 2, forelæsning 12 uge 6.

Så det er et polynomisk basis sæt med M = 3.

Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, software

Automatisk genereret beskrivelseSå kan jeg sørge for least squares ved at sætte mine vægte som:

###### Brug det til at beregne vægtene. ( 20 % )

Vander beregner polynomier af maks orden M-1, og det gør den på horizontantalen .   
Derfor bliver jeg lige nødt til at transponere den, før at den ligning min Phi ud fra ligningerne.

###### Forudsig værdier. Vha de beregnede vægte. Plot punkter og lav en skitse af kurven ved at lave en streg mellem punkter. ( 25 % )

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

Automatisk genereret beskrivelseDe orange punkter er mit test data, og de grønne er de nye punkter, som jeg har prøvet at forudsige.

###### 

###### Givet to nye datapunkter, antag at den originale model [1, x, x^2] ikke producere en acceptabel præcision. Hvad vil jeg så gøre, og hvorfor?

Et billede, der indeholder tekst, skærmbillede, diagram, display/skærm/fremvisning

Automatisk genereret beskrivelseJeg har valgt et polynomium af en grund i starten, så jeg vil fortsætte med at se, om jeg kan fortsætte med det. Jeg vil hæve ordenen og se om det kunne give et acceptabel resultat.   
Jeg ser at en kubic funktion måske vil kunne give den præcision som jeg ønsker. Derfor vil jeg forøge M = 4.   
Derudover, så kunne det også godt ligne en sinus kurve af måske endda lavere orden, så jeg vil måske også begynde at tænke, om det er værd at skifte basis funktion.

#### Tema 2 - Klassifikation

Givet og deres tilhørende klassifikation:   
   
Vist nedenfor.   
Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, linje/række

Automatisk genereret beskrivelse

Brug forskellige klassifikations metoder til at klassificere

###### Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, linje/række Automatisk genereret beskrivelseVha. KNN, K =1.

Distancen til det nærmeste punkt for hver respektive klasse vælger hvilken klasse de nye punkter tager.   
Den klasse med den mindste distance tilhører det nye punkt. Vha. den her metode er det nemt at klassificere T2, ved bare at kigge på plottet.  
   
Men jeg bliver nødt til at regne for T1.   
   
   
Dermed vil klassfikationen vha. KNN være:   
=============  
   
=============



###### Brug NCC til at klassificere punkter.

Et billede, der indeholder tekst, skærmbillede, linje/række, Font/skrifttype

Automatisk genereret beskrivelseEt billede, der indeholder tekst, skærmbillede, linje/række, Font/skrifttype

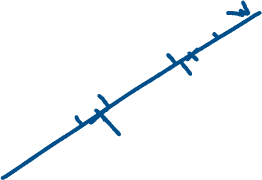
Automatisk genereret beskrivelseDenne metode tager middelværdien af alle dens punkter, og de nye punkters afstande til de her middelværdier, bestemmer hvilken klasse punktet tilhører.



er stadigvæk nem at klassificere, og kan ses fra plot.   
   
Så for bruger jeg beregning.  
C = min (   
   
 )   
Derfor er   
Klassifikationen er så:   
=============  
   
=============

###### Et billede, der indeholder tekst, skærmbillede, display/skærm/fremvisning, software Automatisk genereret beskrivelseEt billede, der indeholder tekst, skærmbillede, linje/række, Font/skrifttype Automatisk genereret beskrivelseBeregn projektions vektoren using the training data points. Ny

Jeg har valgt at lave lineær regression af punkterne til at finde w:



Et billede, der indeholder tekst, skærmbillede, software, Multimediesoftware

Automatisk genereret beskrivelseProjicer data til en 1 dimensional space vha. w.   
Jeg har lavet projicering på w. Og derefter har jeg udplukket alle t værdierne til denne parameterfremstilling.

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

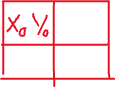
Automatisk genereret beskrivelseEt billede, der indeholder linje/række, diagram, Font/skrifttype, Kurve

Automatisk genereret beskrivelseMiddelværdien af denne, må da være hvor, at min decision boundary skal være, for at adskille klasserne bedst.

Så min decision boundary er ved:   
   
   
   
   
   
   
  
Og med den baggrund, så kan jeg klassificere   
=============  
   
=============

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

Automatisk genereret beskrivelseJeg kunne have valgt en ud fra øjensyn, men jeg gjorde det ud fra gennemsnittet af det 1. dimensionelle data, som så gav mig den mere præcist, vil jeg mene.   
  
  
Min løsning var lidt af en håndværker løsning. Ellers så skulle jeg:   
Fundet   
Ved at maksimere variansen mellem to klasser og minimere variansen   
I klasserne, så kan jeg finde en løsning.  
Beregnet   
 Noget i den stil.

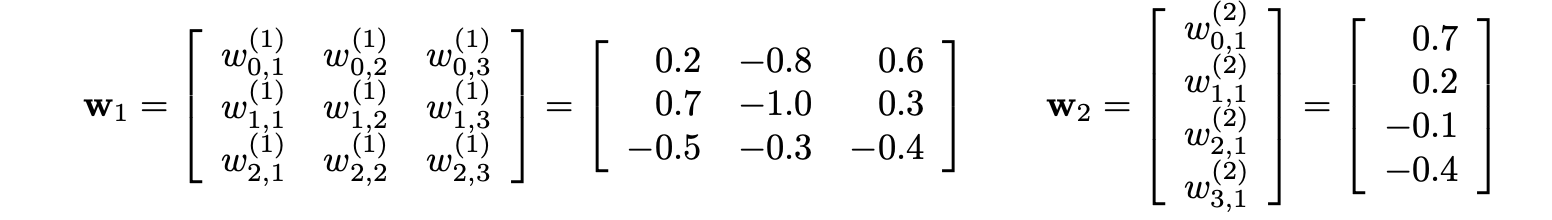
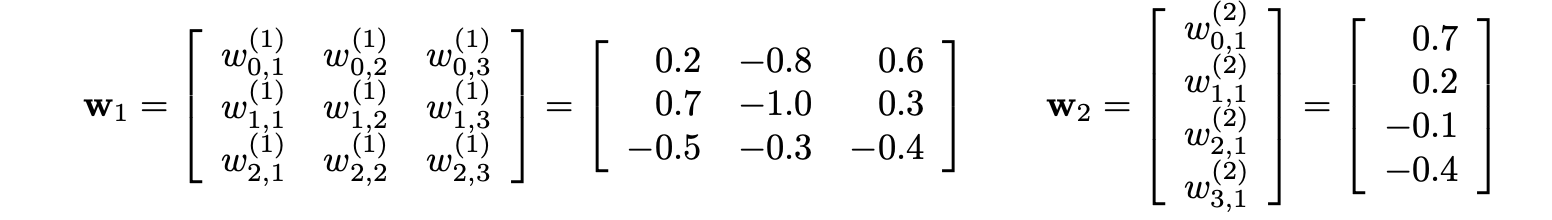


#### Tema 3 - Neural netværk.

Et billede, der indeholder linje/række, cirkel, diagram, skitse

Automatisk genereret beskrivelseEt billede, der indeholder tekst, Font/skrifttype, typografi

Automatisk genereret beskrivelseGivet start vægte og et neuralt netværk som:   
initialiseret til:   
Og et netværk der ser sådan her ud



Et billede, der indeholder tekst, Font/skrifttype, typografi

Automatisk genereret beskrivelse



###### ( 30 % ) Givet et første datapunkt fra træningssættet, hvad er outputtet i det neurale netværk?

Og

###### ( 20 % ) Assuming the output of the network is given the data point . What’s the sum of square errors, if the corresponding target is .

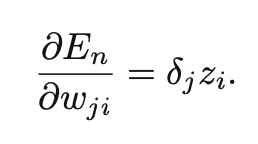
Et billede, der indeholder tekst, Font/skrifttype, hvid, håndskrift

Automatisk genereret beskrivelse

###### Et billede, der indeholder linje/række, cirkel, diagram, skitse Automatisk genereret beskrivelseBack propagation.

Et billede, der indeholder linje/række, cirkel, diagram, Font/skrifttype

Automatisk genereret beskrivelseBeregn

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

Automatisk genereret beskrivelseEt billede, der indeholder Font/skrifttype, hvid, tekst, håndskrift

Automatisk genereret beskrivelse

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

### Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, algebra Automatisk genereret beskrivelseOpgave 1.1 Nyt forsøg. Udledning af w koefficienter ud fra error sum of squares.

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

Automatisk genereret beskrivelse

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

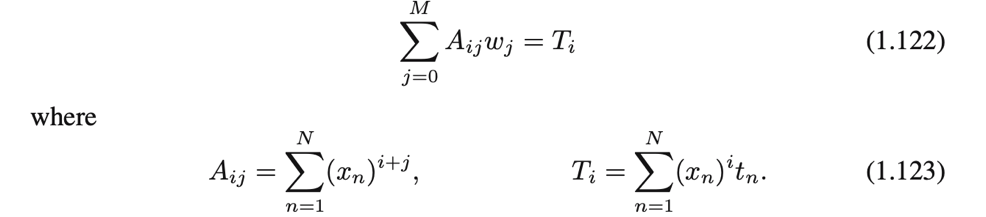
Automatisk genereret beskrivelse

Jeg skal få afledt den, sat den lige med nul for så at finde koefficienterne.   
Jeg afleder i forhold til koefficienterne ,og sætter det til 0, fordi hvorfor ikke?!

Den yderste summering varierer ikke i forhold til .

Der findes en sammenhængen funktion i den variende del.

afhænger af , så så længe at kan jeg ikke sige, at det afledte vil være anderledes fra 0. I det tilfælde at , så gælder der at:

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

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

Hvilket var hvad jeg skulle bevise.

### Opgave 1.1. Gammel forsøg

Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, algebra

Automatisk genereret beskrivelseEt billede, der indeholder tekst, Font/skrifttype, hvid, kalligrafi

Automatisk genereret beskrivelse

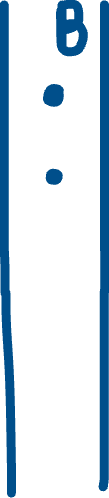
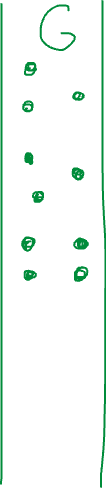
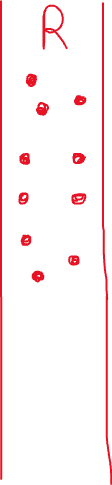
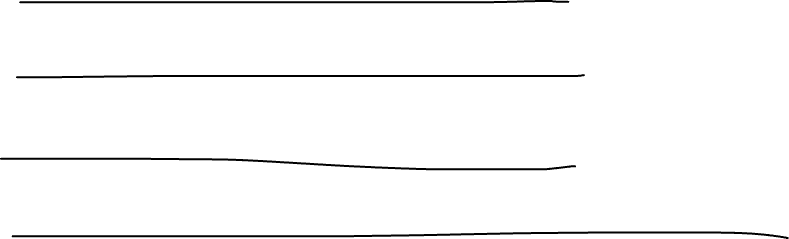
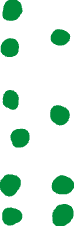
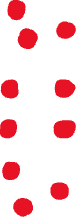
Et billede, der indeholder tekst, Font/skrifttype, hvid, håndskrift

Automatisk genereret beskrivelse

Prøver at minimere.   
Hvis jeg ser på det indre først.

Og så er det, at man kalder det for

### Et billede, der indeholder tekst, Font/skrifttype, skærmbillede, algebra Automatisk genereret beskrivelseOpgave 1.3 Kasser i forskellige farver, hver med forskelligt antal af samme frugter



Hvad hvis jeg siger, at

Sandsynligheden for æbler:

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

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

Det lyder måske meget sandsynligt. Hvis den grønne kun havde haft en frugt, så vil resultatet have set noget anderledes ud.

Så den grønne vægter altså højt, og det er kun fordi, at der cirka er ligevægt i frugter i den grønne, at jeg får så fint et resultat.

*Hvis jeg så skal finde sandsynligheden for, at jeg finder en appelsin i den grønne kasse*.   
Jeg tror jeg ved hvordan jeg skal se på den nu.   
Jeg leder efter sandsynligheden for, kassen jeg trak af var grøn, efter at jeg har trukket en appelsin.

Med bayes theorem kan jeg skrive det om til:

Så nu leder jeg efter sandsynligheden for at trække en appelsin, når jeg trækker i den grønne kasse, sandsynligheden for at kassen som jeg trækker fra er grøn, og sandsynligheden for at trække en appelsin.

Den sidste sandsynlighed er så den mest besværlige:

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

Alternativt kunne jeg have set på at:

Så at sandsynligheden for kassen er grøn, når man trækker en appelsin er det samme som sandsynligheden for, at det er en appelsin man trækker og, at det er en appelsin delt med sandsynligheden for, at det er en appelsin.

### Opgave 1.3 for ny

Jeg har bayes regel:

Og når jeg så bliver spurgt, hvad sandsynligheden for et æble er, så kan jeg beregne for  
, sandsynligheden for at finde et æble, når kassen man trækker fra, er rød.   
 sandsynligheden for at kassen var rød, idet man har trukket et æble.   
  
Den sidste finder jeg ved

### Opgave 1.7 - Den gaussiske distribuering og integral bevis.

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

Automatisk genereret beskrivelseEt billede, der indeholder tekst, skærmbillede, Font/skrifttype, algebra

Automatisk genereret beskrivelse

Jeg ved ikke helt hvad jeg leder efter her, men der er da i hvert fald skridt, som jeg kan lave.  
1. Omskriv til polære koordinator.   
2. Bevis at den gaussiske distribution er normaliseret.

At integrere over alle x til alle y, er det samme som at integrere over alle vinkler til alle længer  
r >= 0. Dermed bliver mine nye grænser:   
Når man skifter fra kartesiske koordinater til polære koordinater, så skal man lige huske at tilføje et r til drdtheta.

2. Skridt  
Jeg skal bruge samme trick med integralet over en faldende eksponentiel til at vise, at den gaussiske funktion ender med at give 1. Det er i samme dur, så jeg hopper videre.

Figur 1: https://en.wikipedia.org/wiki/List\_of\_integrals\_of\_exponential\_functions

### Opgave 1.10

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

Automatisk genereret beskrivelse

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

===========================  
   
===========================  
Det lykkedes mig ikke helt at bevise.

### Opgave 1.12

### Opgave 6.5. Rigtig kernel eller ej? ( Linearitets princip og to ulineære funktioner til ny ulineær funktion )

Givet en kernel funktion, følgende er så også en kernel funktion

Bekræft at resultatet af 6.13 og 6.14 giver rigtige kernels.   
Antaget at jeg har en kernel funktion   
Min nye kernel kan da være:   
   
Som tager den samme form som den valide kernels som jeg havde i starten.

Så kan jeg bare lave en ny ikke lineær, hvor:   
   
   
Og så tager min kernel samme form

### Opgave 6.6. Rigtig kernel eller ej? Polynomial og eksponentielle transformationer.

Kvadratiske og eksponentielle ændringer.

Givet en funktion   
   
Og



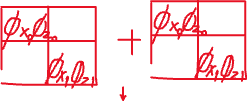
Og så har jeg et biasing led. Jeg kan ikke huske om det præcis var da vi gik fra lineære funktioner til ulineære funktioner, men der justerede vi vores vægte, så der blev tager orden for et bias led. Så det samme kan vi jo også her.   
Vi skal have så at   
   
   
Så laver jeg min nye funktioner så at:

Og igen sørger jeg for dc leden.   
   
   
   
   
Og så har jeg nemlig min kernel på samme form som før.

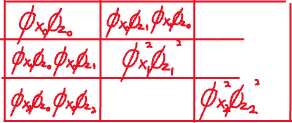
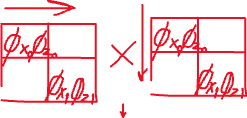
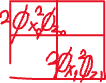
Så for den eksponentielle.   
   
   
   
   
Og så er den på samme form igen.

### Opgave 6.7. Rigtig kernel eller ej? Polynomial og eksponentielle transformationer.

Bevis for 6.17 & 6.18



Så derfor kan jeg bare løse den med:  
   
   
Og så er den af samme form



### Opgave 6.8-9 Rigtig kernel eller ej? Har ikke fortsat

6.19, 6.20

Opgave 6.9   
6.21, 6.21

### Opgave kapitel 6. Kernel PCA

###### Show that kernel PCA with a linear kernel function is equivalent to PCA.

Se på eksempler: Et billede, der indeholder tekst, kvittering, Font/skrifttype, hvid

Automatisk genereret beskrivelseEt billede, der indeholder tekst, kvittering, Font/skrifttype, hvid

Automatisk genereret beskrivelse  
12.71   
12.72

Og så har jeg kernel PCA i formel 12.80  
En faktor her er, at det er linear kernel.   
Det betyder at

Så lad mig udvide:

Hvis jeg sammenligner med:

Så jeg ser, at hvis , så er

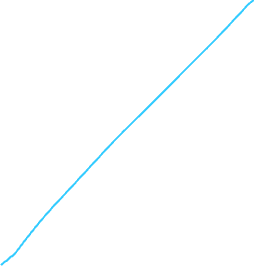
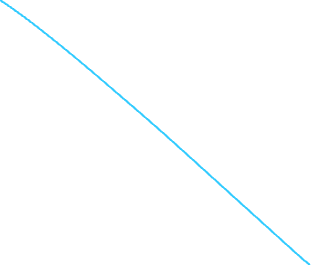
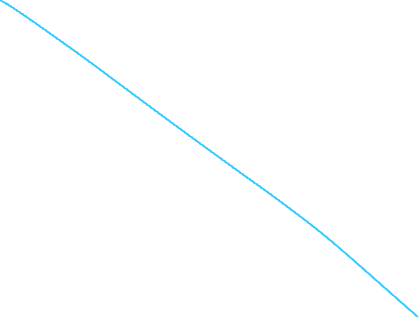
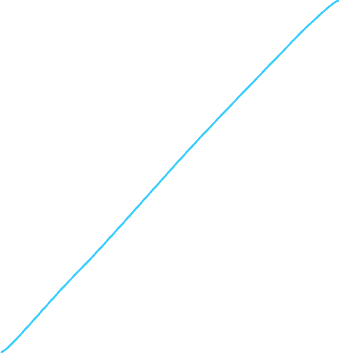
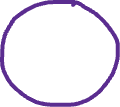
*Men der mangler noget forklaring til, hvorfor egenværdierne skulle være ens. Det er noget med, at op til en hvis dimension, så er de de samme. Det er hvad naveed argumenterer for.*

###### Show that kernel PCA with any kernel function corresponds to PCA applied to arbitrary ( transformed data ) representations of the original data for I = 1, … , N.

### Opgave kapitel 6. Nolinear classification

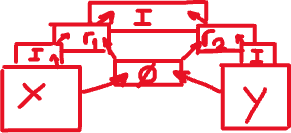
Jeg har et problem på ikke lineær klassifikation, hvor 1. klasse er blå trekanter og 2. klasse er røde cirkler.   
Et billede, der indeholder diagram, linje/række

Automatisk genereret beskrivelse



###### Can you use a Perceptron to classify data from the two classes? If yes, how this can be done? If no, why not?

Med relu, step, sigmoid, identitet kan jeg ikke se, hvordan man skal kunne få det bare med en, men det er jo kernels jeg har om hvor det skulle være muligt, så det kan godt være.   
   
   
   
   
   
Jeg troede, at jeg vha. polære koordinator kunne gøre det, men jeg vil have brug for et netværk bare til at lave vinklen, og så til at beskrive radiusen, og så til at et identitets layer.   
Så det vil blive en   
netværk. Så 2 hidden layers, med 3 perceptrons i alt.   
Hmm det er måske muligt med 1 perceptrons.   
   
Så activationen er



Jeg vil umiddelbart sige nej.

### Opgave 7.2. Show that, if the 1 on the right hand side of the constraint (7.5) is replaced by some arbitrary constant , the solution for the maximum margin hyperplane is unchanged.