Page No.
Date : / /

Correct

Formula used

$$P(w_k|+) = \frac{n_k + 1}{n + |voc|}$$

total of w_k class

total no. of words with repetition allowed

no. of unique words

if a word does not occur in possible case, so, $n_k = 0$

But the whole probablity is not 0, so we are adding 1.

In given case :-

$P(+) = \frac{3}{5}$

$P(I\|+) = \frac{1+1}{14+10} = 0.8333$	$P(I\|-) = \frac{1+1}{6+10}$
$P(the\|+) = \frac{1+1}{14+10} = 0.8333$	$P(the\|-) = \frac{1+1}{6+10}$
$P(a\|+) = \frac{2+1}{14+10}$	$P(a\|-) = \frac{0+1}{6+10}$
$P(activity\|+) = \frac{1+1}{14+10} = 0.8333$	$P(act\|-) = \frac{1+1}{6+10}$
$P(hated\|+) = \frac{0+1}{14+10}$	$P(hated\|-) = \frac{1+1}{6+10}$
$P(loved\|+) = \frac{1+1}{14+10} = 0.8333$	$P(loved\|-) = \frac{0+1}{6+10}$
$P(movie\|+) = \frac{4+1}{14+10}$	$P(movie\|-) = \frac{1+1}{6+10}$
$P(good\|+) = \frac{2+1}{14+10}$	$P(good\|-) = \frac{0+1}{6+10}$

$P(-) = 2/5$

Page No.
Date: / /

$P(great|+) = \frac{2+1}{14+10}$ | $P(great|-) = \frac{0+1}{6+10}$

$P(poor|+) = \frac{0+1}{14+10}$ | $P(poor|-) = \frac{1+1}{6+10}$

Q. I hated the poor acting

(+)

$$\frac{P(I|+) \cdot P(hated|+) \cdot P(the|+) \cdot P(poor|+) \cdot P(acting|+)}{P(+)}$$

$$= \frac{2}{24} \cdot \frac{1}{24} \cdot \frac{2}{24} \cdot \frac{1}{24} \cdot \frac{2}{24} \cdot \frac{3}{5}$$

$$= 6.028 \cdot e^{-7} = 6.028 \times 10^{-7}$$

(−)

$$\frac{P(I|-) \cdot P(hated|-) \cdot P(the|-) \cdot P(poor|-) \cdot P(acting|-)}{P(-)}$$

$$= \frac{2}{16} \cdot \frac{2}{16} \cdot \frac{2}{16} \cdot \frac{2}{16} \cdot \frac{2}{16} \cdot \frac{2}{5}$$

$$= 0.0000122 \qquad 12.2 \times 10^{-7}$$

$\Rightarrow \boxed{P(-|x) > P(+|x)}$

→ belongs to −ve class.

Scanned by CamScanner

Q.	Text	Tag
1	A great game	Sports
2	The election was over	Not Sports
3	Very clean Match	Sports
4	A clean but forgettable game	Sports
5	It was a close election.	Not Sports

a very close game = ?

	A	great	game	The	election	Was	Over	Very	Clean
1.	1	1	1						
2				1	1	1	1		
3								1	1
4	1		1						1
5	1				1	1			

	Match	but	forgettable	It	a	close	Class
1.							Sports
2.							Not Sports
3	1						Sports
4		1	1				Sports
5				1	1		Not Sports.

D → a very close game ?

$P(D|sports) = 2.76 \times 10^{-5}$

$P(D|not\ sports) = 0.572 \times 10^{-5}$

Sports → +, Not sports → −

(for +) $n = 11$, $|voc| = 14$

Page No. 9 (for +)

Date: / /

$P(A|+) = \frac{2+1}{25}$ | $P(A|-) = \frac{1+1}{23}$

$P(great|+) = \frac{1+1}{25}$ | $P(great|-) = \frac{0+1}{23}$

$P(game|+) = \frac{2+1}{25}$ | $P(game|-) = \frac{0+1}{23}$

$P(the|+) = \frac{0+1}{25}$ | $P(the|-) = \frac{1+1}{23}$

$P(election|+) = \frac{0+1}{25}$ | $P(election|-) = \frac{2+1}{23}$

$P(was|+) = \frac{0+1}{25}$ | $P(was|-) = \frac{2+1}{23}$

$P(over|+) = \frac{0+1}{25}$ | $P(over|-) = \frac{1+1}{23}$

$P(very|+) = \frac{1+1}{25}$ | $P(very|-) = \frac{0+1}{23}$

$P(clean|+) = \frac{2+1}{25}$ | $P(clean|-) = \frac{0+1}{23}$

$P(match|+) = \frac{1+1}{25}$ | $P(match|-) = \frac{0+1}{23}$

$P(but|+) = \frac{1+1}{25}$ | $P(but|-) = \frac{0+1}{23}$

$P(forgettable|+) = \frac{1+1}{25}$ | $P(forgettable|-) = \frac{0+1}{23}$

$P(It|+) = \frac{0+1}{25}$ | $P(It|-) = \frac{1+1}{23}$

$P(close|+) = \frac{0+1}{25}$ | $P(close|-) = \frac{1+1}{23}$

$n \to$ a very close game.

Page No. / Date: / /

$P(+|n) = P(a|+) \cdot P(very|+) \cdot P(close|+) \cdot P(game|+) \cdot P(+)$

$= \frac{3}{25} \times \frac{2}{25} \times \frac{1}{25} \times \frac{3}{25} \times \frac{3}{5} = 0.0000276$

$P(-|n) = \frac{2}{23} \times \frac{1}{23} \times \frac{2}{23} \times \frac{1}{23} \times \frac{2}{5} = 0.00000571$

$\Rightarrow P(+|n) > P(-|n)$

$\Rightarrow$ Sports

Q.

Text	Class
Chinese Beijing Chinese	C
Chinese Chinese Sanghai	C
Chinese Macao	C
Tokyo Japan Chinese	J

Chinese Chinese Chinese Tokyo Japan?

$P(C) = 0.0003$

$P(j) = 0.0001$

	Chinese	Beijing	Sanghai	Macao	Tokyo	Japan	Class
1	2	1					C
2	2		1				C
3	1			1			C
4	1				1	1	J

$|voc| = 6$ $\quad P(C) = \frac{3}{4}$, $\quad P(J) = \frac{1}{4}$

n for $(C) = 8$, $\quad n$ for $(J) = 3$

$P(\text{Chinese}/C) = \frac{3+1}{14}$ $\qquad P(\text{Chinese}/J) = \frac{1+1}{9}$

$P(\text{Beijing}/C) = \frac{1+1}{14}$ $\qquad P(\text{Beijling}/J) = \frac{0+1}{9}$

$P(\text{Sanghai}/C) = \frac{1+1}{14}$ $\qquad P(\text{Sanghai}/J) = \frac{0+1}{9}$

$P(\text{Macao}/C) = \frac{1+1}{14}$ $\qquad P(\text{Macao}/J) = \frac{0+1}{9}$

$P(\text{Tokyo}/C) = \frac{0+1}{14}$ $\qquad P(\text{Tokyo}/J) = \frac{1+1}{9}$

$P(\text{Japan}/C) = \frac{0+1}{14}$ $\qquad P(\text{Japan}/J) = \frac{1+1}{9}$

x Chinese Chinese Chinese Tokyo Japan

$P(C/x) = P(\text{Chinese}/C) \times 3 \times P(\text{Tokyo}/C) \times P(\text{Japan}/C) \times P(C)$

$= \frac{4}{14} \times \frac{4}{14} \times \frac{4}{14} \times \frac{1}{14} \times \frac{1}{14} \times \frac{3}{4}$

$= \frac{4}{14} \times \frac{4}{14} \times \frac{4}{14} \times \frac{1}{14} \times \frac{1}{14} \times \frac{3}{4}$

$= \underline{0.000089}$

$P(J/x) = P(\text{Chinese}/J) \times 3 \times P(\text{Tokyo}/J) \times P(\text{Japan}/J) \times P(J)$

$= \left(\frac{2}{9}\right)^3 \times \frac{2}{9} \times \frac{2}{9} \times \frac{1}{4}$

$= \underline{\underline{0.00013}}$

$\Rightarrow P(J/x) > P(C/x)$

Mid Sem Syllabus

1. Naive Bayes
2. Bayesian Classification.
3. Intro to Bayesian decision theory for 2 category Clasf^n, continuous cases, then estimation of posterior probablity
4. Conditional Risk, Expected loss
5. Min Risk Classifier.
6. Min Error Rate Clasf^n
7. Text Clasf^n
8. Text Clasf^n through Naive Bayes.
9. Maximum Likelihood Estimation.
10. Nearest Neighbour (KNN)
11. PDF (Probablity Density f^n).
12. Basic Structure of Pattern
13. Distance measures.

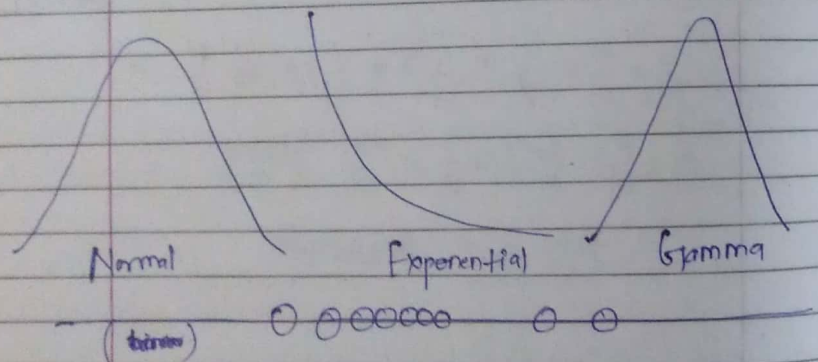
Maximum Likelihood Estimation

→ It is a method of estimating the parameters of statistical models for given observation.

→ The method obtains the parameter estimates by finding the parameter value that maximise the likelihood fn.

→ Maximum likelihood estn is a fn which determines values for the parameters of a model.

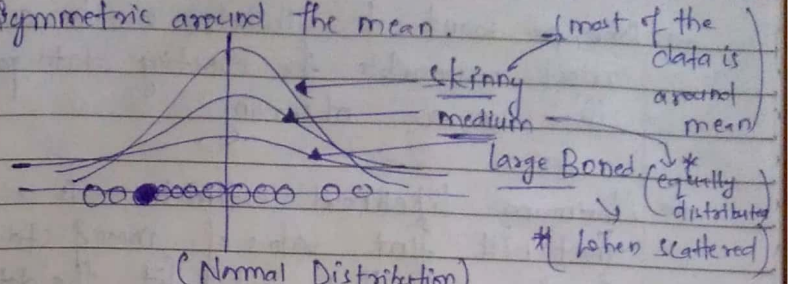
→ The parameter values are found such that; they maximise the likelihood that the process describe by the model produce the data that were actually observed.



Normally Distributed means, at a pt we expect most of the measurements to be close to the mean.

We expect the measurement to be relatively

symmetric around the mean.



(Normal Distribution)

→ For any given or observe data, first we decide which model ~~will~~ we think is the best process. for ~~desirit~~ deciding the data.

→ In this eg. we assume data generation process can be adequately described by gaussian (normal) distribution.

Visual Inspection of the fig suggest that gaussian distribution is a most suitable because most of 10 points are clustered in the middle with few points scattered in the left & right.

→ We know gaussian distribution has 2 parameters :-

(i) mean (μ)
(ii) Standard deviation (σ)

diffn values of these parameters result in diffn curve.

→ Now, we want to calculate, which curve is most reasonable for creating data points that we are observing.

→ Maximum likelihood estimation is a method that will find values of mean & std. deviation for the curve that best fit the data.

For that we want to calculate, if the total probablity of observing all the data i.e. Joint Probablity Distribution of all observed data points which requires calculation of conditional probablities. So, we make assumption, "each data pt. is generated independent of others" for easier computation.

⇒ The probablity density of observing a single data pt. 'x' i.e. generated from a gaussian distribution is given by :,

$$\boxed{P(x ; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}$$

We can figure out the values of mean & std. deviation that results in giving the maximum value of above expression by using calculus given,

(maxima of functions) derivatives

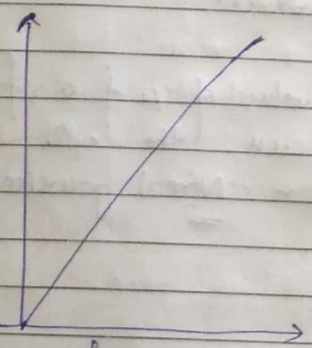
gaussian

The log likelihood :-

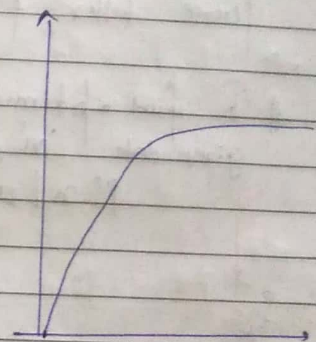
→ The above expression of total probablity is quite difficult to calculate. So, to simplify it we can use natural log of expr.

→ The natural log is, monotonically inc fn. so it is very suitable for this kind of calculation.

→ This is imp bcz it ensures that, the max value of the log of the probablity occured at the same pt. as origin probablity fn.



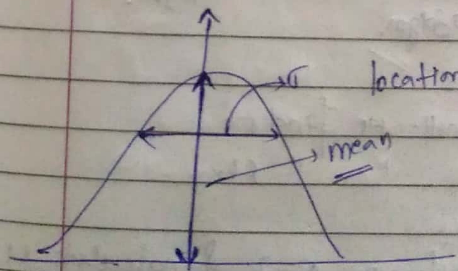
$f(x) = x$



$f(x) = \ln(x)$

Normal gaussian distb

$$P(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

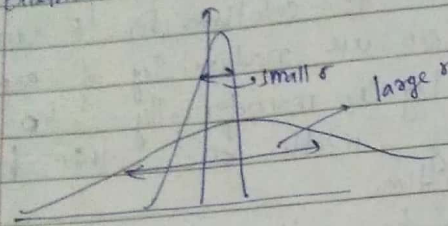


IMP

- μ - determines the location of normal distribution mean.
- if μ is small, curve will shift to left.
- if μ is large, curve will shift to right

Date : / /

- σ - std. deviation that determines normal distribution width.

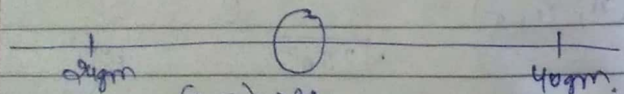


- for larger value of σ, makes the normal curve shorter & wider.
- for smaller value of σ, makes the normal curve taller & narrower.

To find optimal values of μ & σ for given data 'x', we use the eq^n.

$P(x \mid \mu, \sigma) \leftarrow$ Normal gaussian eq^n.

Eg:



$(x=) \rightarrow 32$gm.

by keeping μ constt. we find σ.
" " σ " " " μ.

'σ'=2 (μ)

Let us say, we keep, $\sigma = 2$ & calculate μ?
" " $\mu = 28$ " " σ?

(aim is to find location = 0)
($\sigma = 2$)

Page No.
Date : / /

$$P(32; 28; 2) = \frac{1}{\sqrt{2\pi(2)^2}} e^{-(32-28)^2/2\times(2)^2}$$

$= 0.03$

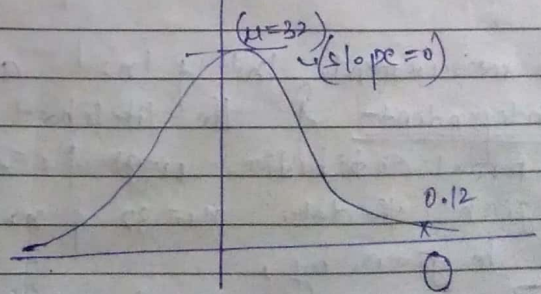
Now, keep, ($\mu = 30$, $\sigma = 2$)

$\boxed{P(32; 30, 2) = 0.12}$

Steps:
- → The y-axis value is likelihood.
- → x-axis is the diff^n values we plugin for μ.

1. Each time we change μ, calculate likelihood & plot it.
2. We can identify the peak in the likelihood graph by determining where, the slope of the curve is 0.

In this eg. (slope = 0, when $\mu = 32$)



Date : / /

* Now, we can fix, $\mu = 32$ & find the optimal value for σ!

Now fix $\mu = 32$ & vary σ for [1, 2, 3 -- till (slope ≠ 0)]

The goal of this eg is to convey the basic concept of, how to find maximum likelihood value for μ & σ.

→ To solve the maximum likelihood estimation for 'μ', we treat 'σ' as a constt & find where, the slope of likelihood fn is '0'.

→ Similarly to solve the maximum likelihood estimation for 'σ', we treat 'μ' as constt & find, when the slope of likelihood fn is '0'.

More than one pts are given.

→ $L_1(\mu = 28, \sigma = 2 \mid x_1 = 32)$

→ $L_2(\mu = 28, \sigma = 2 \mid x_2 = 34)$

$L(\mu = 28, \sigma = 2 \mid x_1 = 32 \text{ \& } x_2 = 34)$

** IMP

assumption in maximum likelihood bcz. if they are dependent we can't find their joint probability.

The measurements 'x_1' & 'x_2' are independent & the likelihood of normal distribution, $\mu = 28$ & $\sigma = 2$ given the data, $x_1 = 32$ & $x_2 = 34$ will be given by, multiplication of L_1 and L_2.

Page No.

Date : / /

$$\Rightarrow \boxed{L = L_1 \times L_2}$$

$$= \frac{1}{\sqrt{2\pi(2)^2}} e^{-(32-28)^2/2\times(2)^2} \times \frac{1}{\sqrt{2\pi(2)^2}} e^{-(34-28)^2/2\times(2)^2}$$

generalising

* for more than one points :-

$$\boxed{L = L_1 \times L_2 \times \cdots \times L_n}$$

→ $L(\mu, \sigma \mid x_1, x_2, \ldots, x_n)$

The overall likelihood fn is the multiplication of all individual likelihood fn.

$$L = L(\mu, \sigma \mid x_1) \times L(\mu, \sigma \mid x_2) \times L(\mu, \sigma \mid x_3) \cdots \times L(\mu, \sigma \mid x_n)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/2\sigma^2} \times \cdots \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2}$$

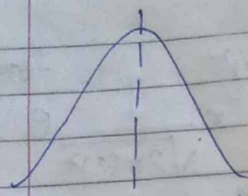
* problem → for finding, μ and σ, we need to maximise whole eqn (1).

IMP

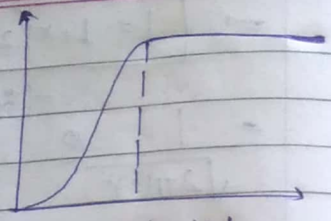
To make calculation simpler, we take log of likelihood fn.

Page No.
Date : / /

Likelihood 3

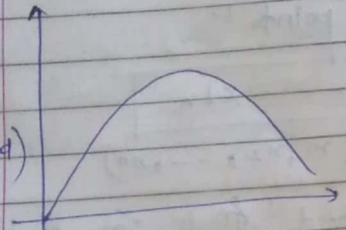


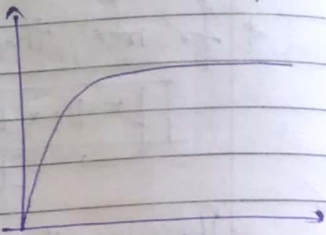
potential value of μ



potential value of σ.

log (likelihood)





$$P(x_i|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for generalized eq^n

$$\ln[L(\mu,\sigma|x_1,x_2,\ldots,x_n)$$

$$=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_1-\mu)^2/2\sigma^2}\times\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_2-\mu)^2/2\sigma^2}\times\cdots\times\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_n-\mu)^2/2\sigma^2}\right)$$

Page No.
Date : / /

$$=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_1-\mu)^2/2\sigma^2}\right)+\cdots+\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x_n-\mu)^2/2\sigma^2}\right)$$

Now, for one term. ($x=x_1$) only.

$$\ln(L)=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-(x_1-\mu)^2/2\sigma^2}\right)$$

$$\ln(L)=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)+\ln\left(e^{-(x_1-\mu)^2/2\sigma^2}\right)$$

$$\ln(L)=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)+\left[\frac{-(x_1-\mu)^2}{2\sigma^2}\right]\ln(e)$$

$$\ln(L)=\ln(2\pi\sigma^2)^{-1/2}-\frac{(x_1-\mu)^2}{2\sigma^2}\underbrace{\ln(e)}_{(1)}\quad(\because \ln e = 1)$$

$$\ln(L)=\frac{-1}{2}\ln(2\pi\sigma^2)-\frac{(x_1-\mu)^2}{2\sigma^2}$$

$$\ln(L)=\frac{-1}{2}\ln(2\pi)-\frac{1}{2}\ln(\sigma^2)-\frac{(x_1-\mu)^2}{2\sigma^2}$$

$$\boxed{\ln(L)=\frac{-1}{2}\ln(2\pi)-\ln(\sigma)-\frac{(x_1-\mu)^2}{2\sigma^2}}$$

for single term
i.e.
$\boxed{x=x_1}$

Page No.
Date : / /

• Generalising $[x = x_1, x_2, \ldots, x_n]$

$\ln(L) = \ln(L(\mu, \sigma / x_1)) + \ldots \ln(L(\mu, \sigma / x_n))$

~~final~~ Now formula:-

$$\ln(L(\mu, \sigma / x_1, x_2, \ldots, x_n)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{(x_1-\mu)^2}{2\sigma^2} - \frac{(x_2-\mu)^2}{2\sigma^2} - \ldots - \frac{(x_n-\mu)^2}{2\sigma^2}$$

I II III

*IMP:-

• We will start by taking derivative wrt 'μ' & use it to find out the peak of the curve log likelihood, where the (slope = 0)

As the I & II term does not contain 'μ', so its derivative would be '0' wrt μ

only the III term, numerator contains the 'μ'

so, eq^n would be

$$\boxed{\frac{d}{d\mu}[\ln(L(\mu, \sigma / x_1, \ldots, x_n)) = \frac{(x_1-\mu)}{\sigma^2} + \ldots + \frac{(x_n-\mu)}{\sigma^2}}$$ *

(Derivative wrt μ) = (slope = 0) → find μ

~~ly we can find 'σ'.

Page No.
Date : / /

$$\boxed{\frac{d}{d\sigma}[\ln(L(\mu, \sigma / x_1, \ldots, x_n))] = -\frac{n}{\sigma} + \frac{(x_1-\mu)^2}{\sigma^3} + \ldots + \frac{(x_n-\mu)^2}{\sigma^3}}$$

↓

Here, put ~~(derivative = 0 i.e. slope = 0)~~ substitute 'μ' from last eq^n & solve for 'σ'.

** This way we find appropriate (μ and σ) for which, log likelihood will be maximised.

* Solving * derivative wrt 'μ'

$$\frac{d}{d\mu}\ln(L(\mu, \sigma ; x_1, x_2, \ldots x_n)) = \frac{1}{\sigma^2}[x_1 + x_2 + \ldots + x_n - n\mu]$$

$$0 = \frac{1}{\sigma^2}[x_1 + x_2 + \ldots + x_n - n\mu]$$

$$0 = x_1 + x_2 + \ldots + x_n - n\mu$$

$$\Rightarrow x_1 + x_2 + \ldots + x_n = n\mu$$

$$\Rightarrow \boxed{\mu = \frac{(x_1 + x_2 + \ldots + x_n)}{n}}$$

$$\frac{d}{d\sigma}\left(\ln(L(\mu,\sigma\,;\,x_1,\dots,x_n)\right)$$

$$= \frac{-n}{\sigma} + \frac{(x_1-\mu)^2}{\sigma^3} + \dots + \frac{(x_n-\mu)^2}{\sigma^3}$$

$$0 = \frac{-n}{\sigma} + \frac{1}{\sigma^3}\left((x_1-\mu)^2 + \dots + (x_n-\mu)^2\right)$$

$$0 = -n + \frac{1}{\sigma^2}\left((x_1-\mu)^2 + \dots + (x_n-\mu)^2\right)$$

$$\Rightarrow \frac{1}{\sigma^2}\left((x_1-\mu)^2 + \dots + (x_n-\mu)^2\right) = n$$

$$\Rightarrow \sigma^2 = \frac{\left((x_1-\mu)^2 + \dots + (x_n-\mu)^2\right)}{n}$$

$$\Rightarrow \boxed{\sigma = \sqrt{\frac{(x_1-\mu)^2 + \dots + (x_n-\mu)^2}{n}}}$$

Syllabus

* Theory of naive bayes & bayesian (self)
  " KNN & maximum likelihood (from notes)