CSC411. GENERATIVE MODELS FOR CLASSIFICATION II.

TWO STAGES:

1) INFERENCE
   How a house should look if good or bad neighbourhood
   Use Bayes Rule for prediction
2) GENERATIVE MODEL
3) FUNCTION CLASSIFICATION

Today: continuous variables.

If a datapt is in more than 1 dimension: MULTIVARIATE.

In any probabilistic model:
   we are calculating the joint distribution.

NAIVE BAYES: All features are independent given that we assume that the house is a good neighbourhood.

This is... weirdly powerful.
· Also, much faster than complete conditionality.

In ML, Gaussians are EVERYWHERE
1) It makes math simple
2) Central Limit Theorem
3) Maximizes data's entropy

} "Why Gaussian?" "What else do u suggest then".

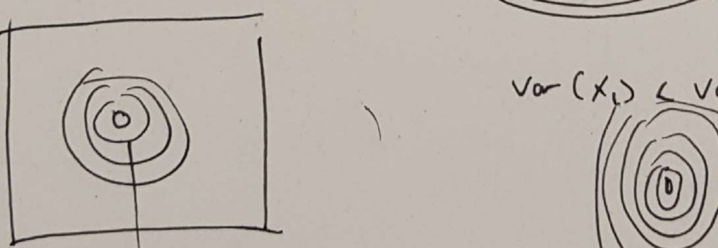
Even if Gaussian distro is a mistaken assumption, a small mistake.

LIKELIHOOD: How my house looks like given a good neighbourhood.
PRIOR, POSTERIOR

d: dimension of data
Σ: covariance of data. (matrix)

Slide 10: assume 2 features
Assume height, weight, and no correlation:

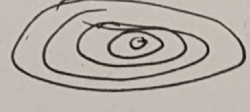
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$



equidistant pts

BUT:

$var(x_1) > var(x_2)$:

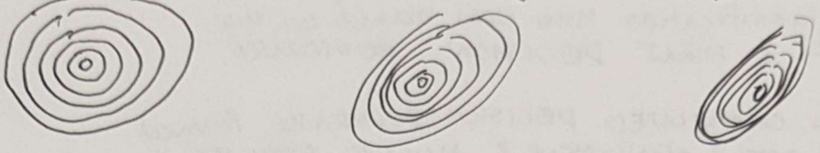


elong. along x_1 dim

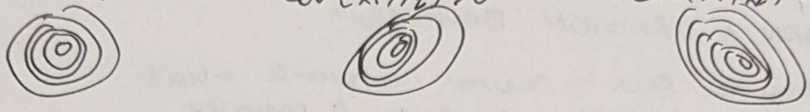
$var(x_1) < var(x_2)$:

elong along x_2 dim.

$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ $\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$



$Cov(x_1, x_2) = 0$ $Cov(x_1, x_2) > 0$ $Cov(x_1, x_2) < 0$



Slide 14: from Bayes rule and took log of both sides.

$$x^T \Sigma_\ell^{-1} x - 2\mu_\ell^T \Sigma_\ell^{-1} x + c \rightarrow (\sigma x^2 - 2\mu\sigma x + c).$$

IF $\Sigma_k = \Sigma_\ell$, decision boundary is linear
IF $\Sigma_k \neq \Sigma_\ell$, decision boundary is quadratic
(happened in logistic regression).

IF Σ is diagonal matrix, we have NAIVE BAYES.
Thus, we only estimate M params for mean. 2n.

$\mathbb{I}[t^{(n)} = 1]$ = indicator fxn: 1 if $t^{(n)} = 1$
0 else

Slide 18: like logistic reg. when assuming shared covariance.
Why do probabilistic inference with this then..?
ANSWER:
In the case of a probabilistic model,
you can get MORE INFO: the PROBABILITY of your house
being in a GOOD or BAD neighbourhood
In logistic: we just get 0 or 1.

Slide 20:
In case of a generative model using Gaussian, even if the boundary is linear, we STILL CAN INJECT OUR PRIOR.

21:
2n parameter

NEURAL NETWORKS.
(Check yeah ans).

So far: all of our classification has been based on the SOLVING OF A NEAT DECISION BOUNDARY.

What if: ... we want a COMPLICATED DECISION BOUNDARY formed SOLELY BY THE DATA? NONLINEAR? MAY NOT EVEN BE A "traditional function"?

ASK THE MODEL TO LEARN A DECISION BOUNDARY?

BASIS FUNCTIONS: can ~~form~~ be fed to neural network that puts them together to form a complex decision ~~boundary~~ boundary we don't know what it looks like.

Some examples: gaussian, sigmoid, polynomial

Slide 8:

w_i = distance btwn this unit and unit before it.

x_i = signal ~~[illegible]~~ from prior unit.

The activation fxn: (the "basis functions")

$$f\left(\sum_i w_i x_i + b\right)$$

Activation fxns must be:
- smooth
- differentiable / gradient nonvanishing

Each unit has its own bias.
TOO MANY HIDDEN UNITS → potential OVERFITTING!

~~NAME~~:

SLIDE 16:
j - index of hidden unit
V_{ji} - weight connecting input i to layer 1 unit j