Lecture 6

* Probabilistic classifier.
* generative model.
* calculates the joint distribution p (X, c), where X is the feature set, c is the class label.
* the Naive Bayes conditional independence assumption, assume that the attribute values are ind
ependent of each other given the class.
* probability values can be calculated quickly. Usually, very fast algorithm.
* Learning : Learning parameters : class prior, likelihood.
  The algorithm learns prior p(c) and likelihood p(xi/c) from the data.
* Prediction / inference : using the likelihood and prior (calculated in learning), the algorithm calculates posterior probability by applying Bayes rule.

x_1	1	0	1	0	1	1	0	0
x_2	0	0	1	1	1	0	1	0
c	1	1	1	0	1	1	0	0

Learning : CPT

c	
1	5/8
0	3/8

$P(x_1/c)$

x_1 \ c	1	0
1	4/5	0/3
0	1/5	3/3
	= 1	= 1

$P(x_2/c)$

x_2 \ c	1	0
1	2/5	2/3
0	3/5	1/3
	= 1	= 1

Inference / Prediction : $x_1=0,\ x_2=1,\ C=?$

$$? \; P(C=1/x_1=0, x_2=1) = \frac{P(x_1=0/c=1)\cdot P(x_2=1/c=1)\cdot P(c=1)}{P(x_1=0, x_2=1)} = (1/5)(2/5)(5/8) = 2/40$$

A

$$? \; P(C=0/x_1=0, x_2=1) = \frac{P(x_1=0/c=0)\cdot P(x_2=1/c=0)\cdot P(c=0)}{P(x_1=0, x_2=1)} = (3/3)(2/3)(3/8) = \frac{1/4}{(2/40)+10}$$

B

$(2/40) + (10/40)$

$$P(x_1=0, x_2=1) = P(x_1=0, x_2=1, c=0) \; (A) \; (B) = 2/40$$
$$+ P(x_1=0, x_2=1, c=1) = 10/40$$

Ex2 : $X_1 = 1, \; X_2 = 1, \; C = ?$

$$P(X_1=1/C=1)\cdot P(X_2=1/C=1)\cdot P(C=1) \;\Big|\; \underbrace{P(X_1=1/C=0)}_{\text{zero}}\cdot P(X_2=1/C=0)\cdot P(C=0)$$

Case I $P(X_i/C) = 0$

Laplace smoothing : Add a small value is numerator & denominator to avoid '0' probability

$$P(X_1=1/C=0) = \frac{\#\,\text{count}\,(X_1=1 \;\&\; C=0) + \varepsilon}{\#\,\text{count}\,(C=0) + 2\varepsilon}$$

Smoothing factor : $\#(X_1 = x_i) / \#(\text{total outcome of } X_i)$

Case II : Missing Data

X_1	1	0	1	0	1	1	0	0
X_2	0	0	1	?	1	0	1	0
C	1	1	1	0	1	1	0	0

Training :- Ignore examples with missing attributes.
→ Replace with mean / most frequent

Prediction :- $P(Y=1/X_1=0)$?

$$= P(X_1=0/Y=1)\cdot P(Y=1)\cdot\Big[\underbrace{P(X_2=0/Y=1) + P(X_2=1/Y=1)}_{1 \,\leftarrow\, \text{Ignore}}\Big]$$

Case III : Continuous Data

x_1	1.2	1.3	1.5	1.6	1.8	1.9	1.4	1.5
x_2	2.1	2.4	2.2	2.6	2.7	2.5	2.8	2.2
C	1	0	1	1	0	0	1	1

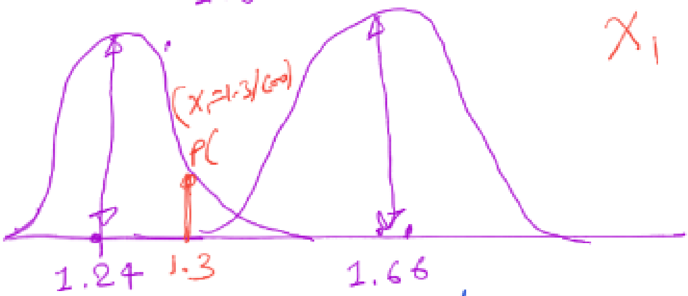
$\mu \quad \sigma$

Assume that x_1 & x_2 are Gaussian.

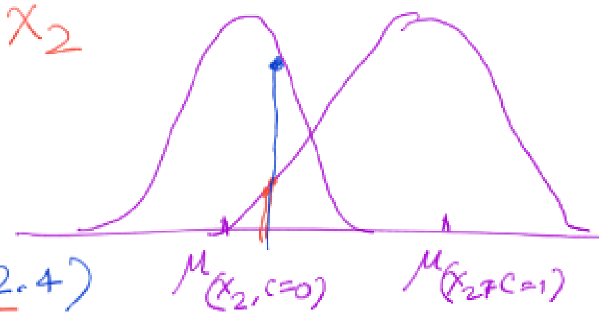
$$\begin{array}{r} 1.8 \\ +\,1.3 \\ +\,1.9 \\ \hline \end{array}$$

$\mu_{(x_1, c=0)}$, $\sigma_{(x_1, c=0)}$

1.66



$\mu_{(x_1 = c = 1)} = 1.24$



$P(Y = 1 / x_1 = 1.3, x_2 = 2.4)$

<u>Case IV</u> Log likelihood

- Probability values can be very small.
- We want to replace multiplication with addition
- Take log of likelihood

$$= \log\left(P(x_i = 1/y=1)\right) + \log\left(P(x_2 = 1/y=1)\right)$$