ECE M146 Discussion 6
Introduction to Machine Learning
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen

1. **Naive Bayes Example** You're stuck in a forest with nothing to eat. Suddently, you spot a mushroom but you don't know if its poisonous. Luckily, you've studied some mushrooms as part of a class to fulfill your undergraduate requirements. Your previous knowledge is summarized by the following chart:

A B C D E

Sample #	IsColorful	IsSmelly	IsSmooth	IsSmall	IsPoisonous
1	0	0	0	1	1
2	0	0	0	0	0
3	1	0	1	1	1
4	1	0	0	0	1
5	0	0	1	0	0
6	0	0	1	0	0
7	1	1	0	0	1
8	1	1	1	0	1
9	0	1	1	0	?

(a) Build a Naive Bayes classifier to classify the unknown mushroom.

Priors $P(E=0) = \frac{3}{8}$

Class Conditional Probabilities

$P(A=0 \mid E=0) = 1$

& $P(A=0 \mid E=1) = \frac{1}{5}$

$P(B=0 \mid E=0) = 1$

$P(B=0 \mid E=1) = \frac{3}{5}$

$P(C=0 \mid E=0) = \frac{1}{3}$

$P(C=0 \mid E=1) = \frac{3}{5}$

$P(D=0 \mid E=0) = 1$

$P(D=0 \mid E=1) = \frac{3}{5}$

Laplace Smoothing

$P(E=0) = \frac{3}{8}$

$P(A=0 \mid E=0) = \frac{3+1}{3+2} = \frac{4}{5}$

$P(A=0 \mid E=0) = \frac{1+1}{5+2} = \frac{2}{7}$

$P(B=0 \mid E=0) = \frac{4}{5}$

$P(B=0 \mid E=1) = \frac{4}{7}$

$P(C=0 \mid E=0) = \frac{2}{5}$

$P(C=0 \mid E=1) = \frac{4}{7}$

$P(D=0 \mid E=0) = \frac{4}{5}$

$P(D=0 \mid E=1) = \frac{4}{7}$

#9	A	B	C	D	E?
	0	1	1	0	

$$P(E=0 \mid A=0, B=1, C=1, D=0) \underset{\hat{E}_9=1}{\overset{\hat{E}_9=0}{\gtrless}} P(E=1 \mid A=0, B=1, C=1, D=0)$$

$$\times P(A=0, B=1, C=1, D=0) \qquad \times P(A=0, B=1, C=1, D=0)$$

$$P(E=0, A=0, B=1, C=1, D=0) \underset{\hat{E}_9=1}{\overset{\hat{E}_9=0}{\gtrless}} P(E=1, A=0, B=1, C=1, D=0)$$

$$P(E=0, A=0, B=1, C=1, D=0)$$

$$= P(E=0)\, P(A=0, B=1, C=1, D=0 \mid E=0)$$

$$= P(E=0)\, P(A=0 \mid E=0)\, P(B=1 \mid E=0)\, P(C=1 \mid E=0)\, P(D=0 \mid E=0)$$

$$= \frac{3}{8} \times 1 \times \boxed{0} \times \frac{2}{3} \times 1 = \boxed{0}$$

$$P(E=1, A=0, B=1, C=1, D=0)$$

$$= \frac{5}{8} \times \frac{1}{5} \times \frac{2}{5} \times \frac{2}{5} \times \frac{3}{5} > 0 \qquad \Rightarrow \hat{E}_9 = 1$$

$$P(E=0, A=0, B=1, C=1, D=0)$$

$$= \frac{3}{8} \times \frac{4}{5} \times \boxed{\frac{1}{5}} \times \frac{3}{5} \times \frac{4}{5} = 0.0288$$

$$P(E=1, A=0, B=1, C=1, D=0)$$

$$= \frac{5}{8} \times \frac{2}{7} \times \frac{3}{7} \times \frac{3}{7} \times \frac{4}{7} = 0.0187 \qquad \Rightarrow \hat{E}_9 = 0$$

(b) Repeat (a) with Laplace smoothing.

2. Discriminative v.s. Generative v.s discriminant function

So far, we have learned two approaches for binary classification in class. The generative approach model the prior $P(C_i)$ and class conditional distribution $P(x|C_i)$. The discriminative approach model $P(C_i|x)$ directly. Taking the Naive Bayes classifier for binary feature and label as an example. It model the class prior as a biased coin and model the class conditional distribution for each feature and each class also as a biased coin. From page 122 of the book *A Course in Machine Learning*, we learned that to make a decision, we can use $\text{sign}(w^T x + b)$ for some w and b. Learning w and b is then a discriminative approach!

(a) What is w and b in terms of the parameters of Naive Bayes classifier?

Generative Approach $\Rightarrow$ Discriminant function

$P(y=0|x) \gtrless P(y=1|x)$ $\boxed{w^Tx+b=0}$?

$P(y=0|x) = P(y=1|x)$ $P(y=0, x) = P(y=1, x)$

(b) Suppose we have D features, how many parameters does the Naive Bayes classifier have? How many parameters does the linear model have?

Naive Bayes : $1 + 2D$

Linear Model : $1 + D$

(c) Can you compare discriminative approach with generative approach given the above example?

Generative Model can generate data

$P(y=0)$ → generate feature x_i based on $P(x_i|y=0)$

$P(y=1)$ →

GAN : Generative Adversarial Network

$$P(y=0, x) = P(y=1, x)$$

$$\log P(y=0, x) - \log P(y=1, x) = 0$$

$$0 = \log\left[\theta_0 \prod_j \theta_{j,0}^{1[x_j=1]} (1-\theta_{j,0})^{1[x_j=0]}\right]$$

$$- \log\left[(1-\theta_0) \prod_j \theta_{j,1}^{1[x_j=1]} (1-\theta_{j,1})^{1[x_j=0]}\right]$$

$$= \log \theta_0 - \log(1-\theta_0) + \sum_j 1[x_j=1]\left(\log \theta_{j,0} - \log \theta_{j,1}\right)$$

$$+ \sum_j 1[x_j=0]\left(\left(\log(1-\theta_{j,0}) - \log(1-\theta_{j,1})\right)\right)$$

$$= \sum_j x_j \log \frac{\theta_{j,0}}{\theta_{j,1}} + \sum_j (1-x_j) \log \frac{1-\theta_{j,0}}{1-\theta_{j,1}} + \log \frac{\theta_0}{1-\theta_0}$$

$$= \sum_j x_j \underbrace{\left[\log \frac{\theta_{j,0}}{\theta_{j,1}} - \log \frac{1-\theta_{j,0}}{1-\theta_{j,1}}\right]}_{w_j} + \underbrace{\sum_j \log \frac{1-\theta_{j,0}}{1-\theta_{j,1}} + \log \frac{\theta_0}{1-\theta_0}}_{b}$$

$$\text{sign}(w^T x + b)$$

3. More about Discriminative v.s. Generative
   Let $p(x|C_1) \sim 0.4\mathbb{N}(0.2, 0.1) + 0.6\mathbb{N}(0.5, 0.1)$. Let $p(x|C_2) \sim \mathbb{N}(0.7, 0.1)$. In MATLAB, plot the two class conditional distribution and find the decision boundary. Let $P(C_1) = P(C_2) = 0.5$, what is the equation to find the posterior distribution for C_1 and C_2. Find and plot the posterior distribution for C_1 and C_2.
   Find the maximum likelihood decision boundary using both the class conditional distribution and the posterior distribution. Comment on your observation.

$$P(x|C_1) \sim 0.4 N(0.2, 0.1) + 0.6 N(0.5, 1)$$

$$P(x|C_2) \sim N(0.7, 0.1)$$

$$P(C_1|x) = \frac{P(C_1, x)}{P(x)} = \frac{P(x|C_1) \times P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

$$P(C_2|x)$$

In Generative Approach,

$$P(x|C_1) \underset{C_2}{\overset{C_1}{\gtrless}} P(x|C_2)$$

In Discriminative Approach,

$$P(C_1|x) \underset{C_2}{\overset{C_1}{\gtrless}} P(C_2|x)$$