Q1-1: Are these statements true or false?
(A) Generative methods model joint probability distribution while discriminative methods model posterior probabilities of Y given X.
(B) We usually train a discriminative model by maximizing the posteriors for true labels for supervised tasks.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false?

(A) Generative methods model joint probability distribution while discriminative methods model posterior probabilities of Y given X.
(B) We usually train a discriminative model by maximizing the posteriors for true labels for supervised tasks.

1. True, True
2. True, False
3. False, True
4. False, False

(A) The aim of a generative model is to learn the generative story, i.e. the joint distribution $P(X, Y)$. On the other hand, a discriminative model aims to directly learn the posterior probability $P(Y \mid X)$.
(B) We usually train a discriminative model by minimizing the corresponding loss function. MLE is also ok, but it often requires us to specify the distribution first, which somehow makes the learning problem more complicated, thus limiting its application area.

Q1-2: Please calculate the MLE λ^{MLE} of Exponential distribution $\lambda e^{-\lambda x}$ for data $D = \{1, 2, 3, 4, 5\}$.

1. 1/15
2. 1/5
3. 1/3
4. 3

Q1-2: Please calculate the MLE λ^{MLE} of Exponential distribution $\lambda e^{-\lambda x}$ for data $D = \{1, 2, 3, 4, 5\}$.

1. 1/15
2. 1/5
3. 1/3
4. 3

By the lecture slides, we have
$\lambda^{MLE} = \frac{N}{\sum_{i=1}^{N} x_i} = \frac{5}{15} = \frac{1}{3}$.

Q2-1: Are these statements true or false?
(A) Naïve Bayes assumes conditional independence of features to decompose the joint probability into the conditional probabilities.
(B) We use the Bayes' rule to calculate the posterior probability.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-1: Are these statements true or false?
(A) Naïve Bayes assumes conditional independence of features to decompose the joint probability into the conditional probabilities.
(B) We use the Bayes' rule to calculate the posterior probability.

1. True, True
2. True, False
3. False, True
4. False, False

(A) Just as we learnt in the lecture.
(B) We use Bayes rule to decompose posterior probability into prior probability and conditional probability given each class, so that we can compute it using the estimated parameters.

Q2-2: Please estimate $P(X, Y)$ for $X = (2, S)$ and $Y = 1$ by the prior probability and conditional probability with the naïve bayes assumptions.
Suppose

$$P(Y = 1) = \frac{5}{10},$$

$$P\left(X^{(1)} = 1 \middle| Y = 1\right) = \frac{2}{5}, P\left(X^{(1)} = 2 \middle| Y = 1\right) = \frac{3}{5},$$

$$P\left(X^{(2)} = S \middle| Y = 1\right) = \frac{1}{5}, P\left(X^{(2)} = M \middle| Y = 1\right) = \frac{2}{5}, P\left(X^{(2)} = L \middle| Y = 1\right) = \frac{2}{5}$$

1. 3/25
2. 1/10
3. 3/50
4. 1/2

Q2-2: Please estimate $P(X,Y)$ for $X=(2,S)$ and $Y=1$ by the prior probability and conditional probability with the naïve bayes assumptions.
Suppose

$$P(Y=1)=\frac{5}{10},$$

$$P\left(X^{(1)}=1 \middle| Y=1\right)=\frac{2}{5}, P\left(X^{(1)}=2 \middle| Y=1\right)=\frac{3}{5},$$

$$P\left(X^{(2)}=S \middle| Y=1\right)=\frac{1}{5}, P\left(X^{(2)}=M \middle| Y=1\right)=\frac{2}{5}, P\left(X^{(2)}=L \middle| Y=1\right)=\frac{2}{5}$$

1. 3/25
2. 1/10
3. 3/50
4. 1/2

$$\mathrm{P}(X=(2,S),Y=1)=P(Y=1)P\left(X^{(1)}=2 \middle| Y=1\right)P\left(X^{(2)}=S \middle| Y=1\right)$$
$$=\frac{5}{10}\times\frac{3}{5}\times\frac{1}{5}=\frac{3}{50}$$

Q3-1: Are these statements true or false?
(A) For instantiations of Naïve Bayes, we directly specify $P(X)$ and $P(Y|X)$, and use MLE to estimate the parameters of these distributions.
(B) We can model $P(Y)$ as multinomial distribution to do multi-class Naïve Bayes.

1. True, True
2. True, False
3. False, True
4. False, False

Q3-1: Are these statements true or false?
(A) For instantiations of Naïve Bayes, we directly specify $P(X)$ and $P(Y|X)$, and use MLE to estimate the parameters of these distributions.
(B) We can model $P(Y)$ as multinomial distribution to do multi-class Naïve Bayes.

1. True, True
2. True, False
3. False, True
4. False, False

(A) In Naïve Bayes, we usually specify $P(Y)$ and $P(X|Y)$ instead of directly modeling the joint distribution.
(B) We can learn it in the lecture slides.

Q3-2: Please estimate $P(X,Y)$ for $X=(2,S)$ and $Y=1$ by Bernoulli Naïve Bayes.

	$X^{(1)}$	$X^{(2)}$	Y
1	1	S	−1
2	1	M	−1
3	1	M	1
4	1	S	1
5	1	S	−1
6	2	S	−1
7	2	M	−1
8	2	M	1
9	2	L	1
10	2	L	1

1. 3/25
2. 1/10
3. 3/50
4. 1/2

Q3-2: Please estimate $P(X,Y)$ for $X=(2,S)$ and $Y=1$ by Bernoulli Naïve Bayes.

	$X^{(1)}$	$X^{(2)}$	Y
1	1	S	-1
2	1	M	-1
3	1	M	1
4	1	S	1
5	1	S	-1
6	2	S	-1
7	2	M	-1
8	2	M	1
9	2	L	1
10	2	L	1

1. 3/25
2. 1/10
3. 3/50
4. 1/2

By the MLE of Bernoulli naïve bayes, we can do the counting.

$$\mathrm{P}(X=(2,S),Y=1)=P(Y=1)P\left(X^{(1)}=2\middle|Y=1\right)P\left(X^{(2)}=S\middle|Y=1\right)$$

$$=\frac{\sum_{i=1}^{N}\mathbb{I}_{y_i=1}}{N}\times\frac{\sum_{i=1}^{N}\mathbb{I}_{x_i^{(1)}=2\,\wedge\,y_i=1}}{\sum_{i=1}^{N}\mathbb{I}_{y_i=1}}\times\frac{\sum_{i=1}^{N}\mathbb{I}_{x_i^{(2)}=S\,\wedge\,y_i=1}}{\sum_{i=1}^{N}\mathbb{I}_{y_i=1}}=\frac{5}{10}\times\frac{3}{5}\times\frac{1}{5}=\frac{3}{50}$$