Naïve Bayes Classifier

Types of Classifiers

- **A classifier is a function that assigns to a sample, $\mathbf{x}$ a class label $\hat{y}$**

$$\hat{y} = f(\mathbf{x})$$

- **A probabilistic classifier obtains conditional distributions $\Pr(Y|\mathbf{x})$, meaning that for a given $\mathbf{x} \in \boldsymbol{X}$, they assign probabilities to all $y \in Y$**
    - Hard classification

$$\hat{y} = \arg\max_{y} \Pr(Y = y|\,\mathbf{x})$$

- **Logistic regression**

$$f(x) = P(Y = 1|\mathbf{x}) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \cdots - \beta_p x_p}}$$

Any other ways to model $\boldsymbol{P}(Y|X)$?

Overview of Naïve Bayes Classifier

- **Probabilistic classifier**
- **Bayes theorem (probability theory)**
- **Strong independence assumption between features**
- **Probability distribution depending on the type of features**
- **Maximum likelihood estimation**

Preliminaries

Bayes Theorem

- **Probability**

$$P(A) = \frac{the\ number\ of\ ways\ that\ event\ A\ occurs}{the\ number\ of\ all\ possible\ outcomes}$$

- Example of sampling with replacement.
  - ✓ The bag includes three blue balls and five white balls

$$P(blue) = \frac{3}{8}$$

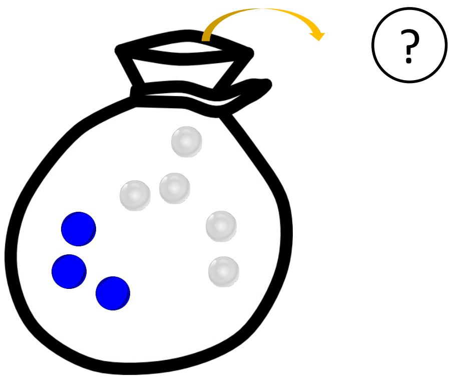
$$P(white) = \frac{5}{8}$$

Bayes Theorem

- **Conditional probability**
  - Suppose a two bags with the same composition.
    - ✓ For independent events A and B,

$$P(A \cap B) = P(A) \cdot P(B)$$
$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

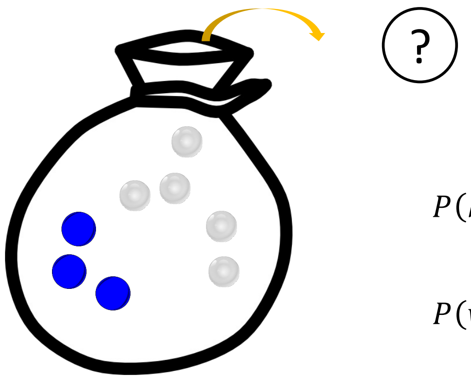
A



$$P(blue) = \frac{3}{8}$$

$$P(white) = \frac{5}{8}$$

B



$$P(blue) = \frac{3}{8}$$

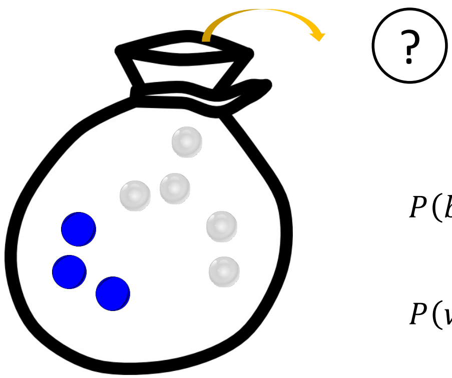
$$P(white) = \frac{5}{8}$$

Bayes Theorem

- **Conditional probability**
  - Suppose a bag with three blue balls and five white balls.
    - ✓ If we draw a ball (A) and subsequently another ball (B) without replacement,

$$P(A \cap B) = P(A) \cdot P(B|A)$$

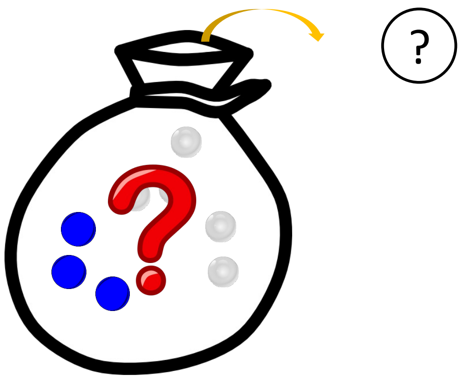
A



$P(blue) = \frac{3}{8}$

$P(white) = \frac{5}{8}$

B



$P(blue) = \frac{?}{7}$

$P(white) = \frac{?}{7}$

Bayes Theorem

- **Conditional probability**
  - Suppose a bag with three blue balls and five white balls.
    - ✓ If we draw a ball (A) and subsequently another ball (B) without replacement,

$$P(A \cap B) = P(A) \cdot P(B|A)$$

A	$P(A)$	B	$P(B\|A)$	$P(A \cap B)$
Blue	$\frac{3}{8}$	Blue	$\frac{2}{7}$	$\frac{3}{8} \times \frac{2}{7} = \frac{6}{56}$
Blue	$\frac{3}{8}$	White	$\frac{5}{7}$	$\frac{3}{8} \times \frac{5}{7} = \frac{15}{56}$
White	$\frac{5}{8}$	Blue	$\frac{3}{7}$	$\frac{5}{8} \times \frac{3}{7} = \frac{15}{56}$
White	$\frac{5}{8}$	White	$\frac{4}{7}$	$\frac{5}{8} \times \frac{4}{7} = \frac{20}{56}$

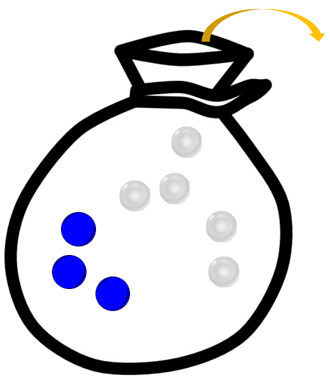
Bayes Theorem

- **Conditional probability**
  - Suppose a bag with three blue balls and five white balls.
    - ✓ If we draw a ball (A) and subsequently another ball (B) without replacement,

$$P(A \cap B) = P(A) \cdot P(B|A)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

A

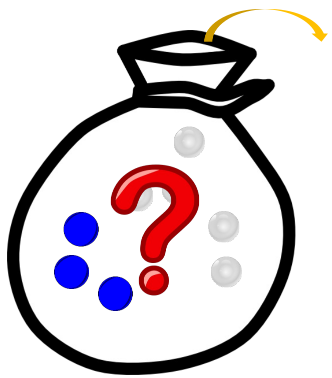


?

$$P(blue) = \frac{3}{8}$$

$$P(white) = \frac{5}{8}$$

B



?

$$P(blue) = \frac{?}{7}$$

$$P(white) = \frac{?}{7}$$

The probability of B depends on the event A

Bayes Theorem

- **Bayes theorem**

$$
\begin{aligned}
&P(A \cap B) = P(B \cap A) \\
&= P(A|B) \cdot P(B) \\
&= P(B|A) \cdot P(A)
\end{aligned}
$$

$$P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

※ Law of Total Probability

- **Total probability**
  - Suppose a bag with three blue balls and five white balls.
    - ✓ If we draw a ball (A) and subsequently another ball (B) without replacement, the probability that the second ball is blue?

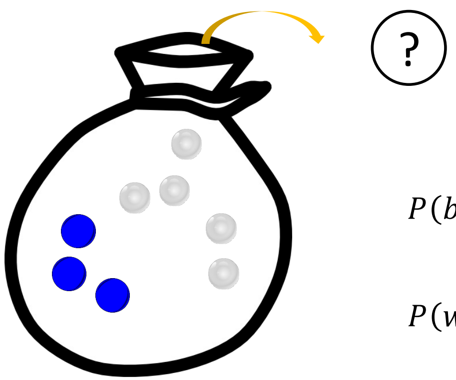
A	$P(A)$	B	$P(B\|A)$	$P(A \cap B)$
Blue	$\frac{3}{8}$	Blue	$\frac{2}{7}$	$\frac{3}{8}\times\frac{2}{7}=\frac{6}{56}$
Blue	$\frac{3}{8}$	White	$\frac{5}{7}$	$\frac{3}{8}\times\frac{5}{7}=\frac{15}{56}$
White	$\frac{5}{8}$	Blue	$\frac{3}{7}$	$\frac{5}{8}\times\frac{3}{7}=\frac{15}{56}$
White	$\frac{5}{8}$	White	$\frac{4}{7}$	$\frac{5}{8}\times\frac{4}{7}=\frac{20}{56}$

$$P(B) = ?$$

- ✓ $P(B = blue) = P(B = blue|A = blue) \cdot P(A = blue) + P(B = blue|A = white) \cdot P(A = white)$
- ✓ $P(B = white) = P(B = white|A = blue) \cdot P(A = blue) + P(B = white|A = white) \cdot P(A = white)$

- ✓ Then, $P(B) = \sum_{a \in S(A)} P(B|A = a) \cdot P(A = a)$, where $S(A)$ is the set of all possible outcomes of A.

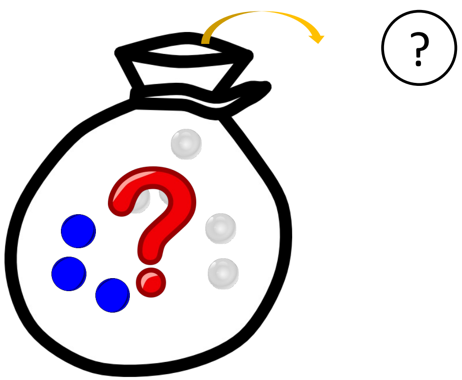
A



$P(blue) = \frac{3}{8}$

$P(white) = \frac{5}{8}$

B

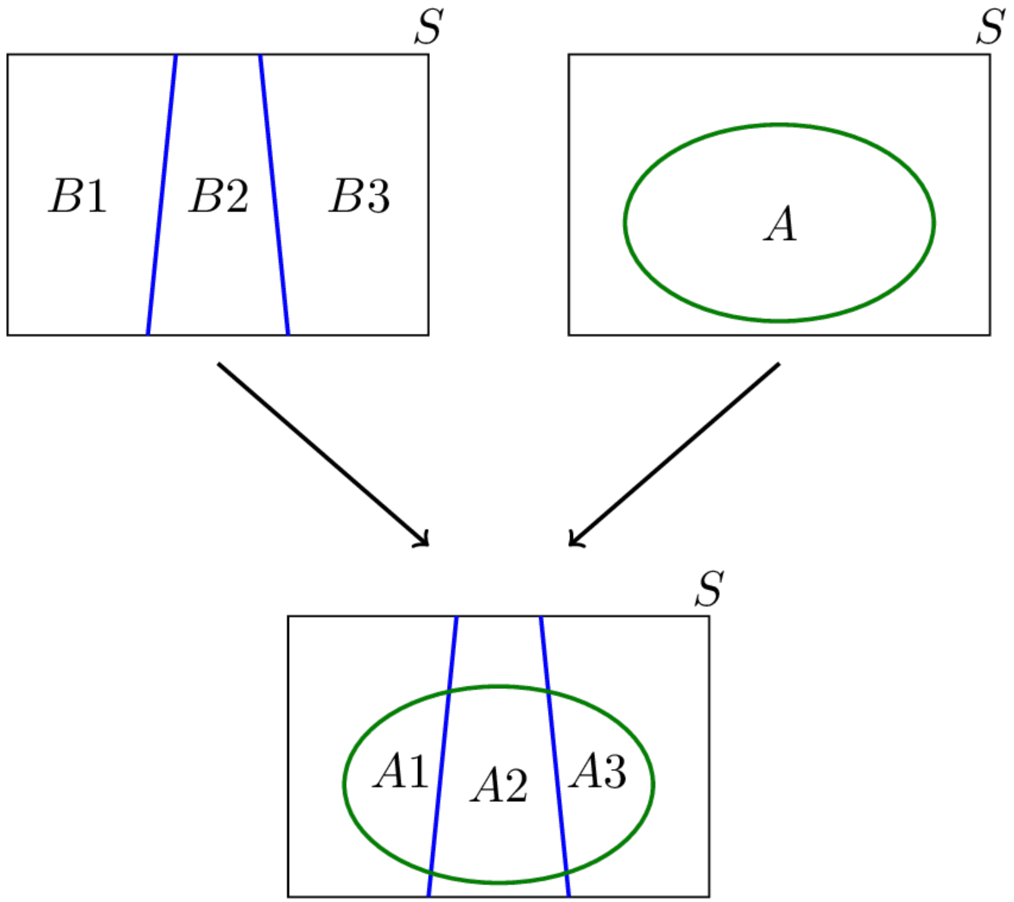


$P(blue) = \frac{?}{7}$

$P(white) = \frac{?}{7}$

※ Law of Total Probability

$$P(A) = \sum_{b \in B} P(b)P(A|b)$$



Motivating Example
for Naïve Bayes Classifier

Classification Example

- **Prior knowledge**
  - Suppose a classifier that predicts whether the person is man or woman
  - If we do not know anything about the person, how can you make a prediction?
    - ✓ $P(male) = 0.5$
    - ✓ $P(female) = 0.5$

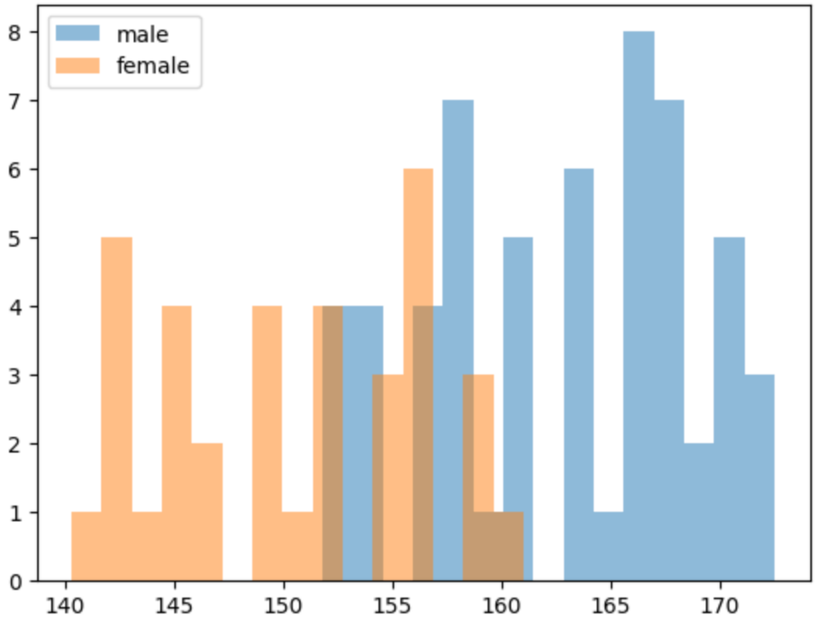
  - What if we restrict the population to engineering college?
    - ✓ $P(male) > p(female)$

- **Prior probability**
  - Without exact knowledge prior probability, it can be estimated as the ratio in the training data.
  - E.g., the ratio of male and female in our training dataset.

Classification Example

- **Likelihood from other features**
  - Now suppose that the height of the given person is 167cm.
  - We will use the conditional probability such as
    - ✓ $P(male|167cm)$ and $P(female|167cm)$

Estimated probability density



We can extract information from the training dataset.

Classification Example

- **Final prediction = Prior knowledge + Likelihood from features**
  - For final prediction, we should integrate the prior knowledge and likelihood from features.

  - Suppose an extremely male-dominant population with prior probability
    - ✓ $P(male) = 0.99$
    - ✓ $P(female) = 0.01$
  - If we know that the height of the given person is 150cm, how would we predict their gender?
    - ✓ Despite the high likelihood, we may still find it difficult to predict that the person is female.

Due to the prior probability!

Classification Example

- **Final prediction = Prior knowledge + Likelihood from features**
  - For final prediction, we should integrate the prior knowledge and likelihood from features
  - Generally, we set the height is known as h
    - ✓ The likelihood that we will predict that the person is male
      - $P(male) \times P(height = h|male)$
    - ✓ The likelihood that we will predict that the person is female
      - $P(female) \times P(height = h|female)$
  - If we have additional information about the weight w, we assume conditional independence between features for simplicity.
    - ✓ $P(male) \times P(height = h|male) \times P(weight = w|male)$
    - ✓ $P(female) \times P(height = h|female) \times P(weight = w|female)$

Conditional independence assumption is not True in real-world!

  - Based on which likelihood is higher, we will make the prediction whether the person is male or female.

Naïve Bayes Classifier

Naïve Bayes Classifier

- **Conditional probability model for Naïve Bayes classifier**
  - Naïve Bayes classifier calculates following probability for every class

$$p\left(C_k \middle| x_1, \dots, x_p\right) = p(C_k|\mathbf{x})$$

✓ x_i represents each feature (independent variable)

✓ k represents k-th class and classifier assigns output class with the maximum probability

  - Re-formulation using Bayes' theorem

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

✓ This equation is also written as

$$\text{posterior} = \frac{\text{prior}\times\text{likelihood}}{\text{evidence}}$$

Naïve Bayes Classifier

- **Naïve Bayes Classifier**

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

  - Denominator $p(\mathbf{x})$ does not depend on class

$$\underset{k}{\operatorname{argmax}}\, p(C_k|\mathbf{x}) = \underset{k}{\operatorname{argmax}}\, p(C_k)p(\mathbf{x}|C_k)$$

- $p(C_k)p(\mathbf{x}|C_k)$ **is equivalent to the joint probability** $p(C_k, x_1, \dots, x_p)$
  - Using chain rule $p(C_k, x_1, \dots, x_p)$ can be written as follows

$$\begin{aligned} p(C_k, x_1, \dots, x_p) &= p(C_k)p(x_1, \dots, x_p|C_k) = p(C_k)p(x_1|C_k)p(x_2, \dots, x_p|C_k, x_1) \\ &= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1) \cdots p(x_p|C_k, x_1, \dots, x_{p-1}) \end{aligned}$$

  - Naïve Bayes classifier assumes conditional independence of each feature

$$p(x_i|C_k, x_j) = p(x_i|C_k)$$
$$p(x_i|C_k, x_j, x_l) = p(x_i|C_k)$$

※ Chain Rule

- **Chain rule permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities**

$$P(A_n, \ldots, A_1) = P(A_n | A_{n-1}, \ldots, A_1) \cdot P(A_{n-1}, \ldots, A_1)$$

- Repeating this process with each final term creates the product

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k | \bigcap_{j=1}^{k-1} A_j\right)$$

Naïve Bayes Classifier

- **Naïve Bayes Classifier**

$$p(C_k)p(x_1|C_k)p(x_2|C_k,x_1)\cdots p\left(x_p\middle|C_k,x_1,\ldots,x_{p-1}\right)$$

$$= p(C_k)p(x_1|C_k)p(x_2|C_k)\cdots p\left(x_p\middle|C_k\right) = p(C_k)\prod_{i=1}^{p} p(x_i|C_k)$$

$$\therefore p(C_k)p(\mathbf{x}|C_k) = p(C_k)\prod_{i=1}^{p} p(x_i|C_k)$$

- **Decision function of Naïve Bayes classifier**

$$\hat{y} = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}} p(C_k|\mathbf{x}) = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}} p(C_k)\prod_{i=1}^{p} p(x_i|C_k)$$

Example of Parameter Estimation for Naïve Bayes Classifier

Spam Mail Classification

- **Spam e-mail classification**
  - Given an e-mail, we need to decide whether it is spam or not.
  - How?
    - ✓ We first need to make a feature vector $\boldsymbol{x}$ from the e-mail texts.
    - ✓ The number of words in the world = N_v
    - ✓ We make a binary vector $x \in \{0, 1\}^{N_v}$ where each element indicates the existence of corresponding word.
    - ✓ Some words more frequently occur in spam mails.
      - "mortgage", "loan", "buy", "replica"
    - ✓ Some words are neutral
      - "the", "a", "some", "is"

Spam Mail Classification

- **Suppose that the number of words N_v is 10,000 for simplicity**
  - For prior probability $P(y)$, we can use Bernoulli distribution (binary classification)
  - From the naïve bayes, $P(x_1, \ldots, x_{10000}|y) = P(x_1|y) \cdots P(x_{10000}|y)$
    - ✓ For each $P(x_i|y)$, two distribution $P(x_i|y=0)$ and $P(x_i|y=1)$ should be estimated.
    - ✓ Since we have used binary feature for all x_i, we can use Bernoulli distribution.
      - $P(x_i|y=0) = P(x_i=1|y=0)^{x_i} \cdot P(x_i=0|y=0)^{(1-x_i)}$
      - $P(x_i|y=1) = P(x_i=1|y=1)^{x_i} \cdot P(x_i=0|y=1)^{(1-x_i)}$
  - Now we have probability distribution for all elements, we can calculate the likelihood of the dataset (N samples).
    - ✓ $L = \prod_{n=1}^{N} p(y_n) \cdot \prod_{i=1}^{10000} p\,(x_i|y_n)$
    - ✓ We can expand the likelihood equation using Bernoulli distribution or calculate the log likelihood instead, and then maximize it to find MLE estimators for all parameters.

Parameter Estimation
for Naïve Bayes Classifier

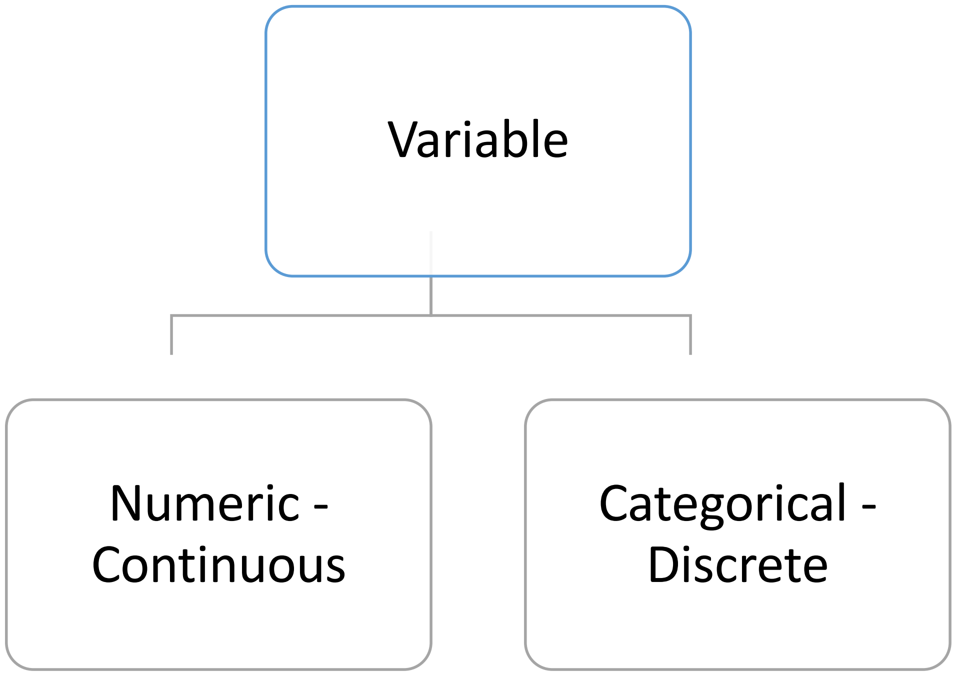
Naïve Bayes Classifier

$$p(C_k)\prod_{i=1}^{p} p(x_i|C_k)$$

- **Parameter estimation and event models**
  - A class' prior setting: by assuming equi-probable classes ($p(C_k) = 1/K$) or by calculating an estimate for the class probability from the training set ($p(C_k) = n_k/n$)
  - Select appropriate probability distribution for $p(x_i|C_k)$
    - ✓ For continuous variables, Gaussian distribution is the common choice
    - ✓ For discrete variables, multinomial distribution is the common choice
  - After setting, probabilistic model for naïve Bayes classifier, parameters of distributions are estimated using training data
    - ✓ Calculate $\tilde{p}\left(C_{y_i}\middle|\mathbf{x}\right) = (C_{y_j})\prod_{i=1}^{p} p(x_i|C_{y_j})$ for j-th sample
    - ✓ Calculate $\tilde{p}(\boldsymbol{C}|\mathbf{X}) = \prod_{j=1}^{n} \tilde{p}\left(C_{y_i}\middle|\mathbf{x}\right)$ and maximize this probability

Naïve Bayes Classifier

- **Determine** $p(x_i|C_k)$
  - The probability functions depend on the type of variables
    - ✓ Probability distributions for discrete random variables: Binomial, Multinomial, Geometric, …
    - ✓ Probability distributions for continuous random variables: Gaussian(Normal), χ^2, beta, F, t, …



Naïve Bayes Classifier: Discrete

- **Bernoulli naïve Bayes**
  - In the multivariate Bernoulli event model, features are independent booleans (binary variables) describing inputs
    - ✓ Ex) Each variable only takes values 0 or 1
  - Each x_i is a boolean expressing the occurrence of event

$$p(\mathbf{x}|C_k) = \prod_{i=1}^{d} p_{ki}^{x_i} (1 - p_{ki})^{1-x_i}$$

    - ✓ d is the number of input features
    - ✓ p_{ki} is the probability that x_i is 1(true) for class k

※ Bernoulli distribution

- **The probability distribution of a random variable which takes the value 1 with success probability of p and the value 0 with failure probability of $q = 1 - p$**
  - For random variable following Bernoulli distribution,
$$p(X = 1) = 1 - p(X = 0) = p = 1 - q$$
  - Probability mass function over possible outcomes y
$$f(y; p) = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{if } y = 0 \end{cases}$$
    - ✓ This can also be expressed as
$$f(y; p) = p^y (1 - p)^{1-y} \quad \text{for } y \in \{1,0\}$$
  - Expected value of a Bernoulli random variable X
$$\mathbb{E}[X] = p$$
  - Variance of a Bernoulli random variable
$$\mathrm{Var}[X] = p(1 - p)$$

Naïve Bayes Classifier: Estimation of Parameters

- **Maximum likelihood estimates for Bernoulli Naïve Bayes model**
  - Each input variable can take values 0 or 1
  - There exist n samples with d features
  - Prior probability $p(C_k)$

$$p(C_k) = \frac{n_k}{n}$$

✓ n_k is the number of samples that y_i belongs to class k

  - Likelihood $p(\mathbf{x}|C_k)$

$$p(\mathbf{x}|C_k) = \prod_{i=1}^{d} p_{ki}^{x_i}(1 - p_{ki})^{1-x_i}$$

  - Posterior probability $p(C_k|\mathbf{x})$

$$p(C_k|\mathbf{x}) \propto p(C_k)p(\mathbf{x}|C_k)$$

Naïve Bayes Classifier: Estimation of Parameters

- **Maximum likelihood estimates for Bernoulli Naïve Bayes model**

$$L = \prod_{j=1}^{n} p\left(C_{y_j}\right) p\left(\mathbf{x}_j \middle| C_{y_j}\right) = \prod_{j=1}^{n} \frac{n_{y_j}}{n} \left(\prod_{i=1}^{d} p_{y_j i}^{x_{ji}} \left(1 - p_{y_j i}\right)^{1 - x_{ji}} \right)$$

$$\log L = \sum_{j=1}^{n} \log \frac{n_{y_j}}{n} + \sum_{j=1}^{n} \sum_{i=1}^{d} (x_{ji} \log p_{y_j i} + \left(1 - x_{ji}\right) \log(1 - p_{y_j i}))$$

  - ✓ n is the total number of data samples
  - ✓ n_{y_j} is the number of data samples belong to class y_j ($y_j \in \{1,2, \dots k\}$)
  - ✓ x_{ji} is the i-th input variable's value for j-th data sample

- Parameters to be estimates
  - ✓ For each class k, probability to occur 1 for each feature i, p_{ki}

Naïve Bayes Classifier: Estimation of Parameters

- **Maximum likelihood estimates for Bernoulli Naïve Bayes model**
  - To obtain optimal p_{ki}, set $\frac{\partial \log L}{\partial p_{ki}} = 0$

$$\frac{\partial \log L}{\partial p_{ki}} = \sum_{j \in \{m: y_m = k\}} \left\{ \frac{x_{ji}}{p_{ki}} - \frac{1 - x_{ji}}{1 - p_{ki}} \right\}$$

$$= \frac{|\{m: x_{mi} = 1, y_m = k\}|}{p_{ki}} - \frac{|\{m: x_{mi} = 0, y_m = k\}|}{1 - p_{ki}} = 0$$

✓ $|\{m: x_{mi} = 1, y_m = k\}|$ is the number of data samples in set of $\{m: x_{mi} = 1, y_m = k\}$

✓ $\{m: x_{mi} = 1, y_m = k\}$ is a set that contains every sample with $x_i = 1$ in class k

✓ $|\{m: x_{mi} = 1, y_m = k\}|$+$|\{m: x_{mi} = 0, y_m = k\}|$=$n_k$

$$p_{ki} = \frac{|\{m: x_{mi} = 1, y_m = k\}|}{n_k}$$

Naïve Bayes Classifier: Estimation of Parameters

$$\log L = \sum_{j=1}^{n} \log \frac{n_{y_j}}{n} + \sum_{j=1}^{n} \sum_{i=1}^{d} (x_{ji} \log p_{y_j i} + \left(1 - x_{ji}\right) \log(1 - p_{y_j i}))$$

Naïve Bayes Classifier: Estimation of Parameters

x	y
1	0
1	0
1	0
0	0
0	1
0	1
1	1

$$p(x = 0|y = 0) = p_{00} = \frac{1}{4}$$

$$p(x = 1|y = 0) = p_{01} = \frac{3}{4}$$

$$p(x = 0|y = 1) = p_{10} = \frac{2}{3}$$

$$p(x = 1|y = 1) = p_{11} = \frac{1}{3}$$

Naïve Bayes Classifier: Discrete

- **Discrete random variables with more than two outcomes**

x	y
High	0
Mid	0
High	0
Low	0
High	0
Low	0
High	0
High	0

$$p(x = Hihg|y = 0) = p_{0,High} = \frac{5}{8}$$

$$p(x = Mid|y = 0) = p_{0,Mid} = \frac{1}{8}$$

$$p(x = Low|y = 0) = p_{0,Low} = \frac{2}{8}$$

Naïve Bayes Classifier: Discrete

- **Multinomial naïve Bayes**
  - With a multinomial event model, samples represent the frequencies with which certain events have been generated by a multinomial $(p_1, \ldots, p_d)$
    - ✓ p_i is the probability that event i occurs
    - ✓ m is the number of features in input data

$$p(\mathbf{x}|C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{k_i}^{x_i}$$

    - ✓ $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ represents each sample and it can be seen as a histogram with x_i counting the number of times event i was observed in a particular instance
  - This is the event model typically used for document classification
    - ✓ With events representing the occurrence of a word in a single document
      → bag of words representation

※ Multinomial distribution

- **Multinomial distribution is a generalization of the binomial distribution**
  - Binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments with success probability p

$$p(X = k) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

  - Example of binomial distribution is the distribution of the number of head when flipping a coin n times (in this case, $p = 0.5$)
    - ✓ Probability that k times head occur among n trials

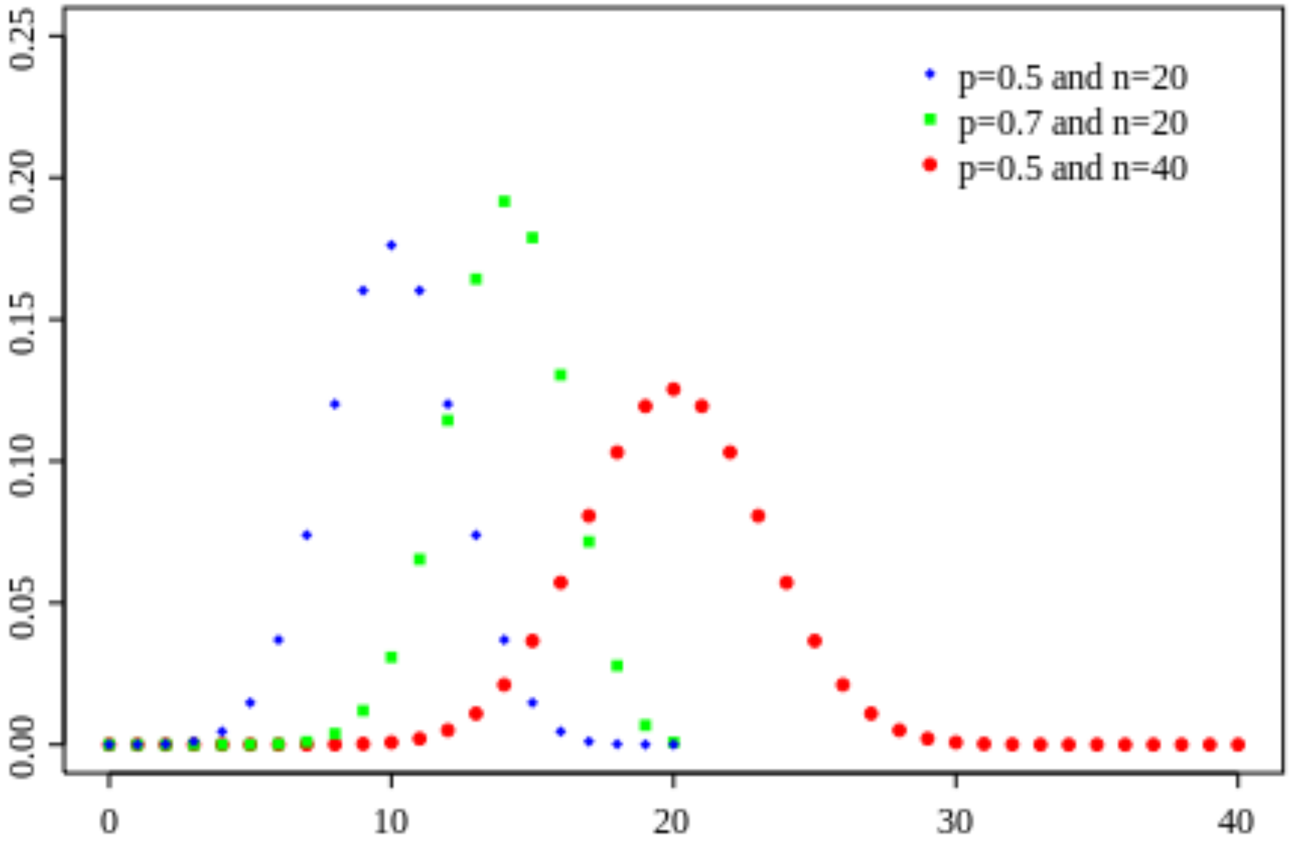
$$p(X = k) = \frac{n!}{k!\,(n-k)!} 0.5^k 0.5^{n-k} = \frac{n!}{k!\,(n-k)!} 0.5^n$$

  - In multinomial distribution, possible outcome is more than two and each outcome has its own probability to occur, $(p_1, \dots, p_d)$
    - ✓ $p_1 + \cdots + p_d = 1$
    - ✓ d is the number of possible outcomes
    - ✓ $n_{\mathbf{x}} = \sum_{i=1}^{d} x_i$

$$p(\mathbf{x} = (x_1, x_2, \dots, x_d)) = \frac{n_{\mathbf{x}}!}{x_1! \cdots x_d!} p_1^{x_1} \cdots p_d^{x_d}$$

※ Multinomial distribution

- **Binomial distribution**



The Application of Naïve Bayes

- **Text Classification**
  - Spam detection
  - Authorship identification
  - Age/gender identification
  - Language identification

- **Input**
  - A document d

- **Output**
  - A fixed set of classes $C = \{C_1, C_2, \ldots, C_K\}$

Naïve Bayes Classifier: Discrete

- **Multinomial naïve Bayes**
  - The multinomial naive Bayes classifier becomes a linear classifier when expressed in log-space

$$\log p(C_k | \mathbf{x}) \propto \log p(C_k) \prod_i p_{k_i}^{x_i} = \log p(C_k) + \sum_{i=1}^{n} x_i \log p_{ki} = b + \mathbf{w}_k^T \mathbf{x}$$

✓ $b = \log p(C_k)$

✓ $w_{ki} = \log p_{ki}$

✓ $\frac{(\sum_i x_i)!}{\prod_i x_i!}$ term only depends on $\mathbf{x}$ and does not depend on class

Naïve Bayes Classifier

- **Gaussian naïve Bayes**
  - When dealing with continuous data, a typical assumption is that the continuous values associated with each class are distributed according to a Gaussian distribution

$$p(x = v|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-u_k)^2}{2\sigma_k^2}}$$

$$p(C_k) \prod_{i=1}^{p} p(x_i|C_k) = p(C_k) \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} e^{-\frac{(v_i-u_{ki})^2}{2\sigma_{ki}^2}}$$

Thank you!

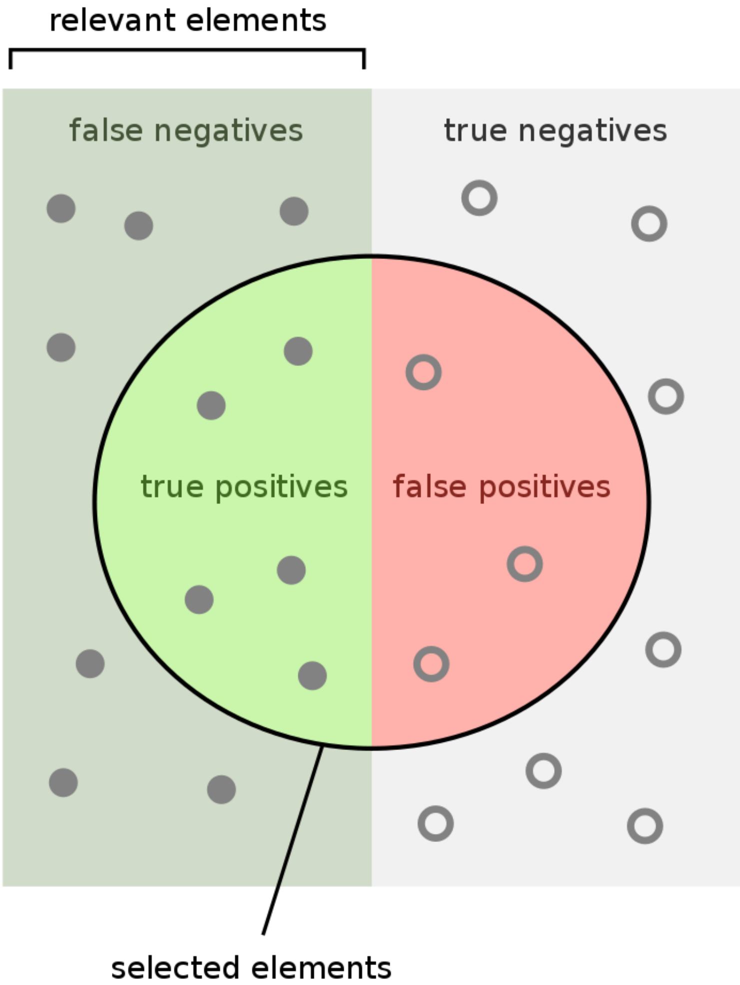
Example of Bayes' Theorem

- **Suppose a drug test is 99% sensitive and 99% specific**
  - 99% sensitive=99% true positive over real positive
  - 99% specific=99% true negative over real negative

Real / Decision	Positive	Negative
Positive	True positive	False positive (Type Ⅰ error)
Negative	False negative (Type Ⅱ error)	True negative

- **Suppose that 0.5% of people are users of the drug**

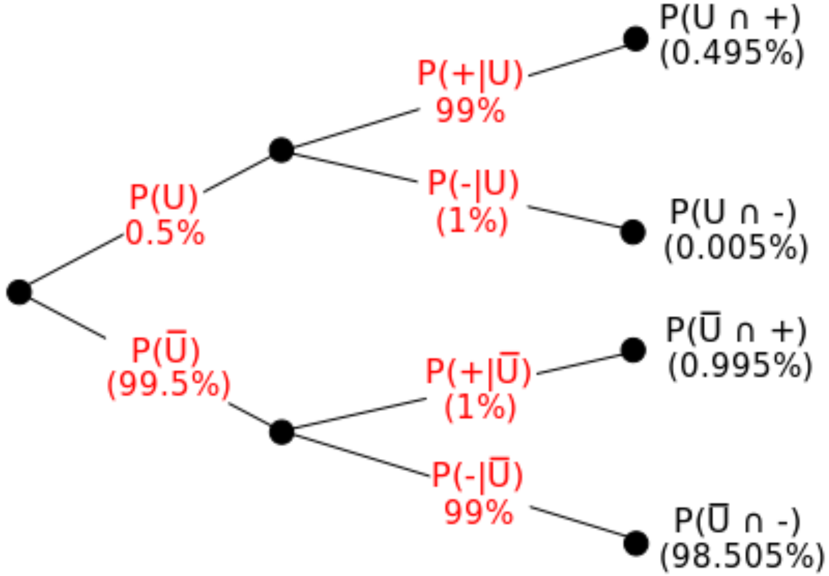
※ Sensitivity and Specificity



<https://en.wikipedia.org/wiki/Sensitivity_and_specificity>

Example of Bayes' Theorem

- **If a randomly selected individual tests positive, what is the probability he or she is a user of drug?**
  - This problem is to calculate $P(U|+)$
    - ✓ $+$ means positive drug test
    - ✓ U represents user, $\overline{U}$ represents non-user



$$P(U|+) = \frac{P(U)P(+|U)}{P(+)} = \frac{P(U)P(+|U)}{P(U)P(+|U) + P(\overline{U})P(+|\overline{U})} = \frac{0.005 \times 0.99}{0.005 \times 0.99 + 0.995 \times 0.001}$$

$$\approx 33.2\%$$

※ $P(A) = \sum_{b \in B} P(b)P(A|b)$