Mobile Robot
Probability and Bayesian Classifier
Lecture 5

Jeong-Yean Yang

2020/10/22

Dept. of Intelligent Robot Eng. MU

Probability

- Probability
  - Pr(x) = 0.111…
- Sum of all possibilities.

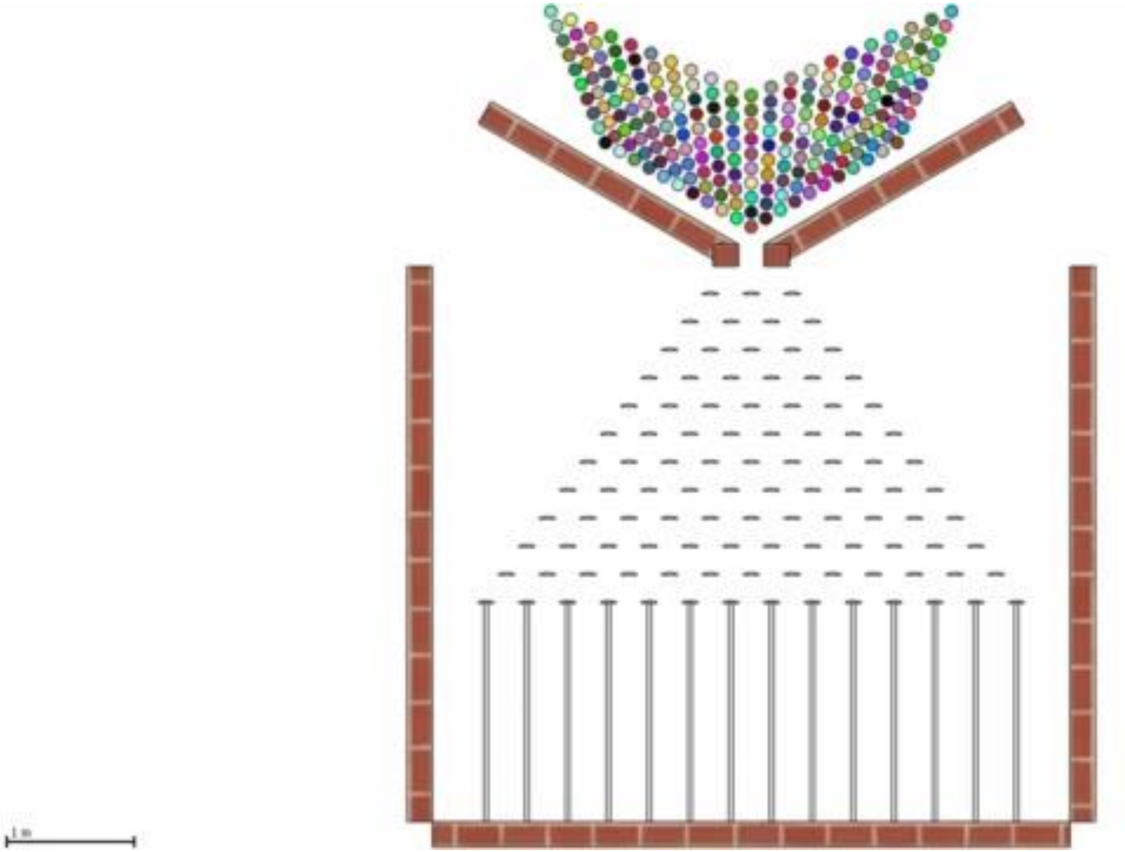
$$\sum \Pr(x) = 1$$

- Continuous domain

$$\int \Pr(x)dx = 1$$

- You already learned about probability..
  - Korean education is so tough….T_T….

Gaussian Probability Generation

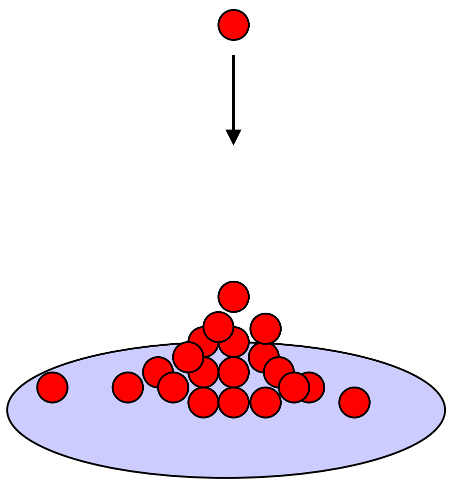


$$\Pr(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right)$$

With C++ or Python, How to Generate Gaussian Distribution?

- Rand() returns integer from 0 to RAND_MAX(32767)
  - Rand() is NOT Gaussian(Normal) distribution

- Remind the video



*Marsaglia polar method

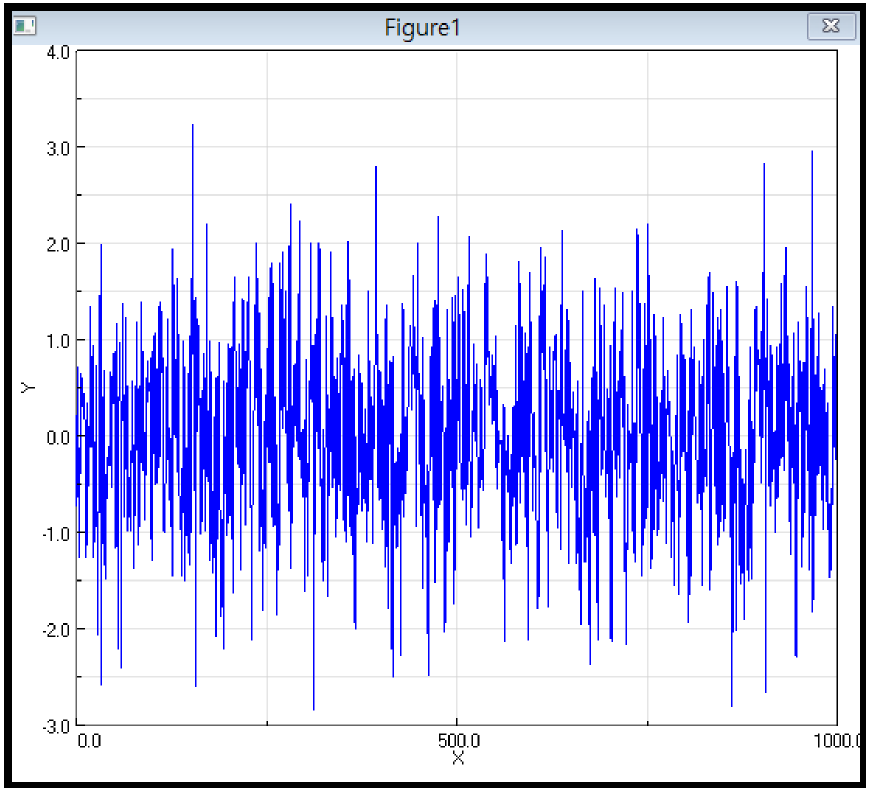
$$r \sim N(0,1)$$

```
double u,v,r;
while(1)
{
    u=2*rand()/((double)RAND_MAX)-1;
    v=2*rand()/((double)RAND_MAX)-1;

    r=u*u+v*v;
    if (r==0 || r>1)    continue;
    break;
}

r   = sqrt(-2*log(r)/r);
r   = u*r;
```

N(0,1) returns Gaussian Distribution



1000 samples

randn(1,1000) generates 1000 samples

Question:

How we generate x with mean and standard deviation?

$$x \sim N(0,1)$$

$$x' \sim N(\mu, \sigma)?$$

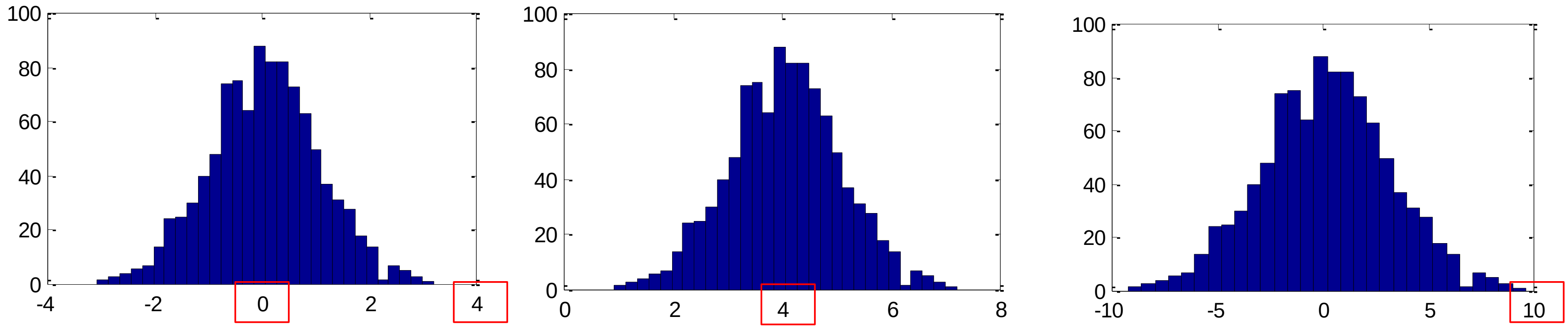
Gaussian Generation $x' \sim N(\mu, \sigma)$

- Mean value: μ is a offset from 0

$$x \sim N(0,1) \quad \Rightarrow \quad x' \sim N(0,1) + \mu = N(\mu,1)$$

- Standard deviation

$$x \sim N(0,1) \quad \Rightarrow \quad x' \sim \sigma N(0,1) = N(0,\sigma)$$



$$x \sim N(0,1)$$

$$x' = x + 4$$

$$x' \sim N(0,1) + 4 = N(4,1)$$

$$x' = 3x$$

$$x' \sim 3N(0,1) = N(0,3)$$

Gaussian Distribution or Normal Distribution(Z)

$$z \sim N(0,1) \qquad z = \frac{x-\mu}{\sigma} \sim N(0,1)$$

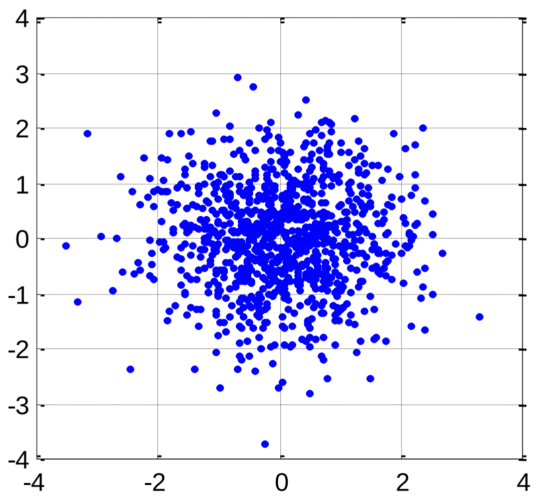
$$x \sim \sigma N(0,1) + \mu = N(\mu, \sigma)$$

- We learn it at high school, TT.
- Z is called "Normal Distribution" $\Pr(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right)$
- X is normalized with mean and standard deviation

$$\Pr(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Probability in 2D Space

- How to generate 2D Gaussian Prob.?
  - Easy. A= randn(1000,2) and plot(A(:,1),A(:,2),'.')

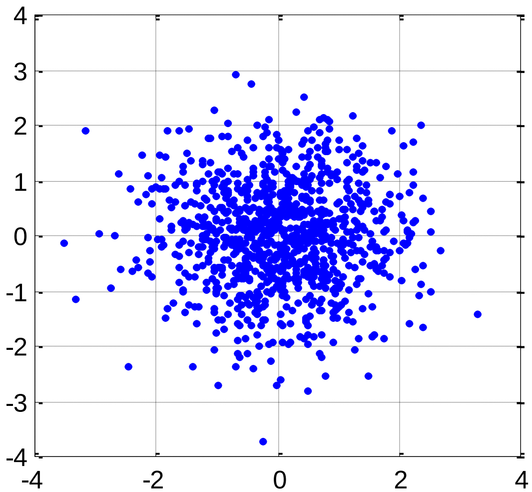


Plot(A(:,1),A(:,2),'.')

1 DIM $z_1 \sim N(0,1)$

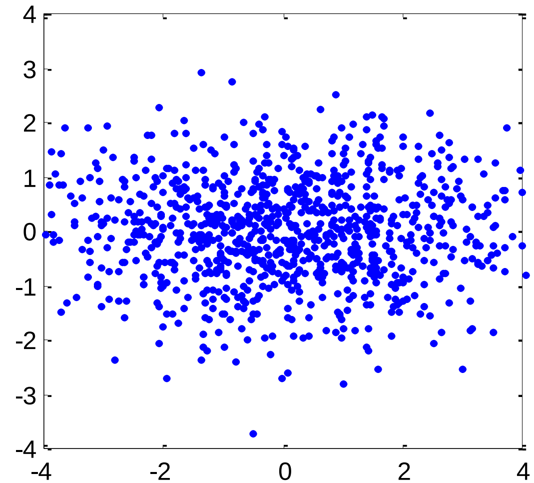
2 DIM $z_2 = \begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma\right)$

$\mu = \begin{pmatrix} x_{mean} \\ y_{mean} \end{pmatrix}$ $\sigma = ?$



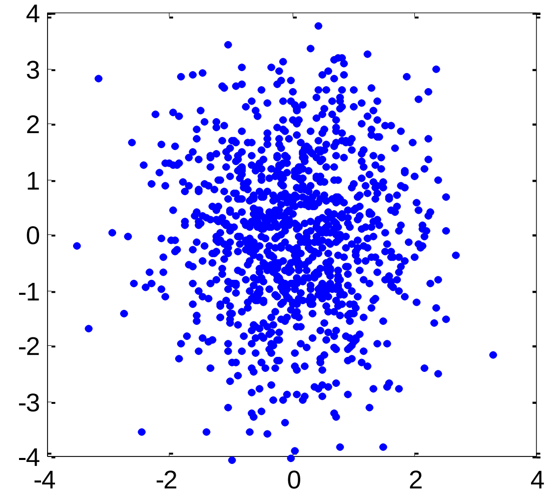
Plot(A(:,1),A(:,2),'.')

$$z_2 = \begin{pmatrix} x \\ y \end{pmatrix}$$



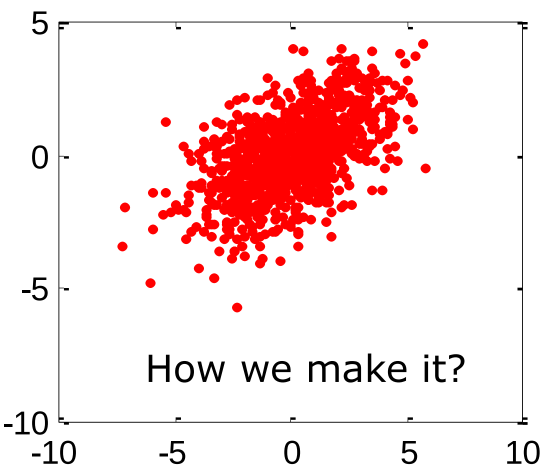
Plot(2*A(:,1),A(:,2),'.')

$$z'_2 = \begin{pmatrix} 2x \\ y \end{pmatrix}$$



Plot(A(:,1), 1.5*A(:,2),'.')

$$z'_2 = \begin{pmatrix} x \\ 1.5y \end{pmatrix}$$



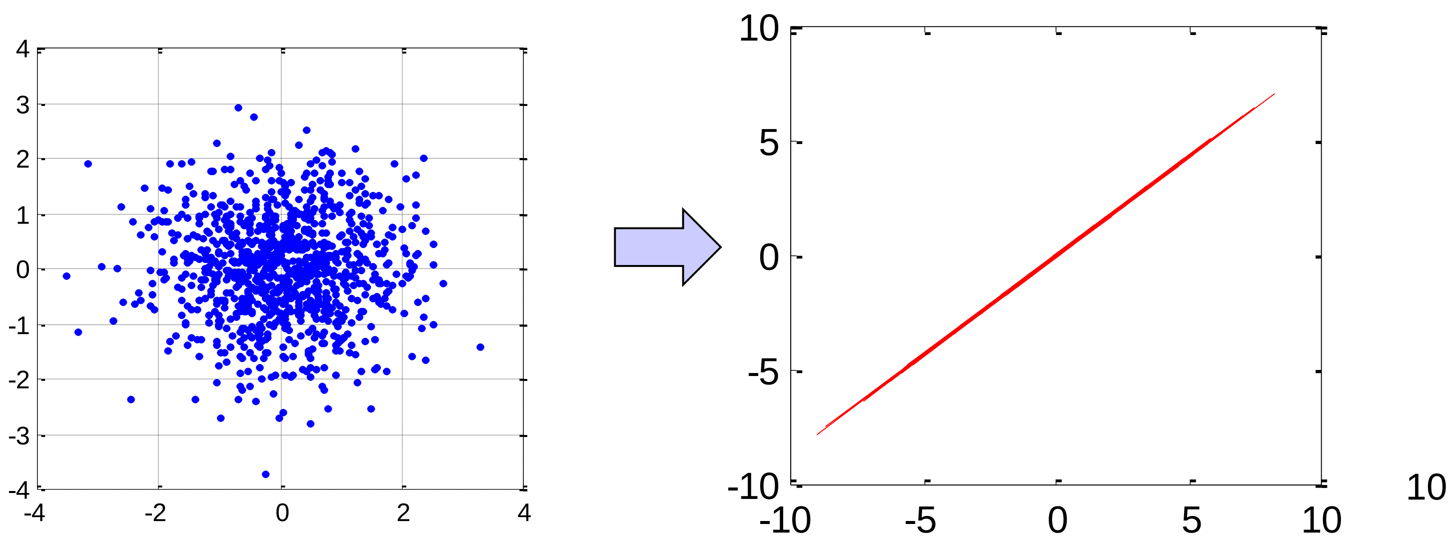
$$z' = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1.5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \Sigma \begin{pmatrix} x \\ y \end{pmatrix}$$

Quiz

$$z' = \begin{pmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 1.5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

How it will distribute?

$$\text{Hint} : Det \begin{pmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 1.5 \end{pmatrix} = 3 - 3 = 0$$



Probability in n-dim. Space

- 1Dim

$$\Pr(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \qquad x \sim N(\mu, \sigma)$$

- N-Dim

$$\Pr(\hat{x}) = \left(Det(2\pi\Sigma)\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right) \qquad \hat{x} \sim N(\hat{\mu}, \Sigma)$$

- Look, Sigma matrix

$$\Sigma = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1.5 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1.5 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \dots & 0.5 \\ 0.5 & \dots \end{pmatrix}$$

Scale factor for principal axis

Rotation

Important for Map matching

Two types of Probability

- A Priori Probability
  - When you use probability, you use a prior probability

$$\Pr(A) = 0.6$$

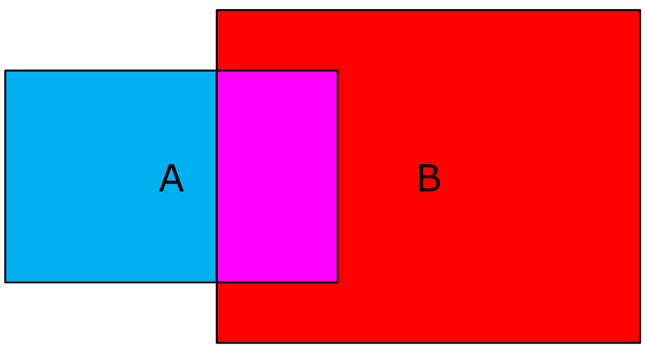
- Posterior Probability (Conditional probability)
  - Bayesian probability
  - Prob. Of A on condition that B occurs,

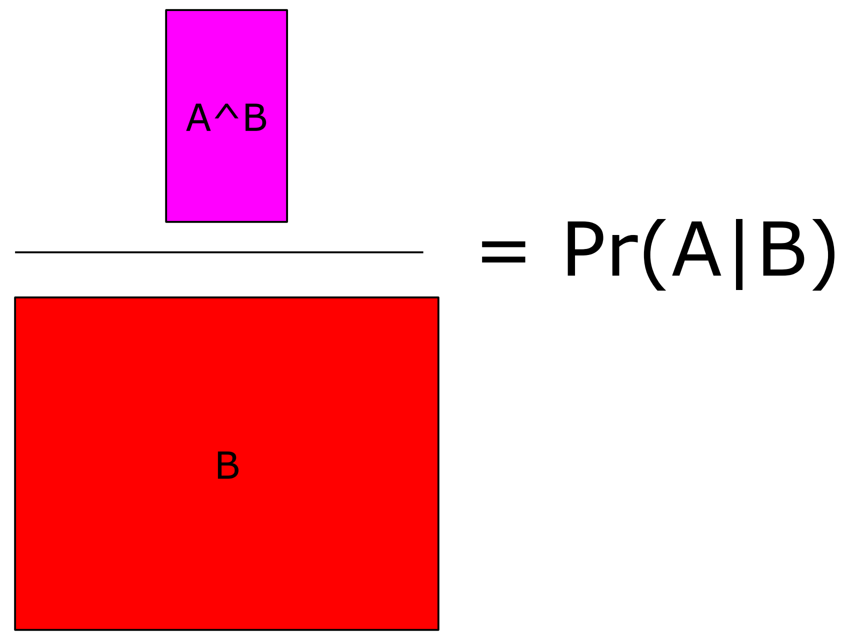
$$\Pr(A \mid B) = 0.6$$

- A prior and Posterior probability are very different.

Conditional Probability

- What is Pr(A|B)?
  - Probability of A under the Probability of B
  - Or Probability of A within the given B





Why Posterior Prob. Is very different?

- Rock-Paper-Scissors game.
  - Prob(Rock) = 1/3

- When a player did “Rock” before,
  - Prob(Rock) is still 1/3? -→ No, in general.

Dept. of Intelligent Robot Eng. MU

Posterior Prob.

- When events A and B occur,
- P(A): Probability of A occurrence
- P(B): Probability of B occurrence.
- P(A^B): Probability of Both A and B occurrence
- Definition:

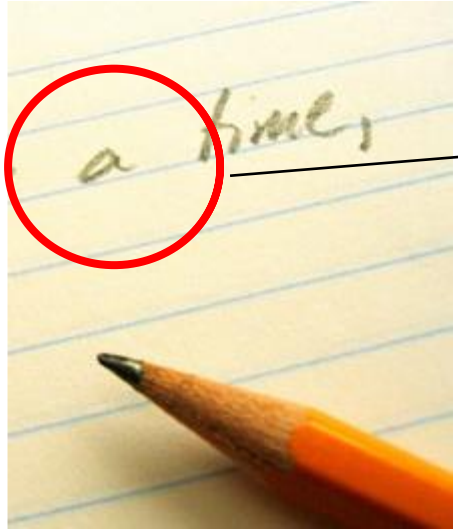
$$\therefore \Pr(A|B) = \frac{P(\mathrm{A}\wedge \mathrm{B})}{P(B)}$$

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

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

Why Posterior Probability?
It reduces Classification Errors..

- What is Classification?



- When a data x is given, is it a specific class, C?
  - It is called, "classification"

$$if \;\; x \in C \; or \; Not$$

Is It Big or Not?

<

- Normal human can say that..
  - Right is bigger.. ^_^..

- When a X is given, can you say that "it is big or not"?

X

Posterior Prob.

- When events A and B occur,
- P(A): Probability of A occurrence
- P(B): Probability of B occurrence.
- P(A^B): Probability of Both A and B occurrence
- Definition:

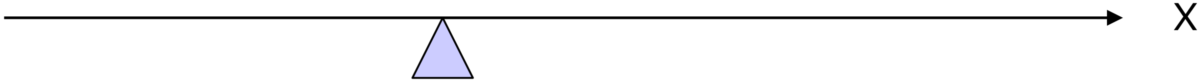
$$\therefore \Pr(A \mid B) = \frac{P(\mathrm{A} \wedge \mathrm{B})}{P(B)}$$

$$P(A \mid B)P(B) = P(A \wedge B) = P(B \mid A)P(A)$$

$$\therefore \Pr(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Classification: Bayesian Classifier

- Random variable, x : probability of event occurrence.
- When x is given, is x involved in class w1 or w2?
  - Ex)
  - Assume X is height,
  - When x = 170, is it tall(w1) or not(w2) ?



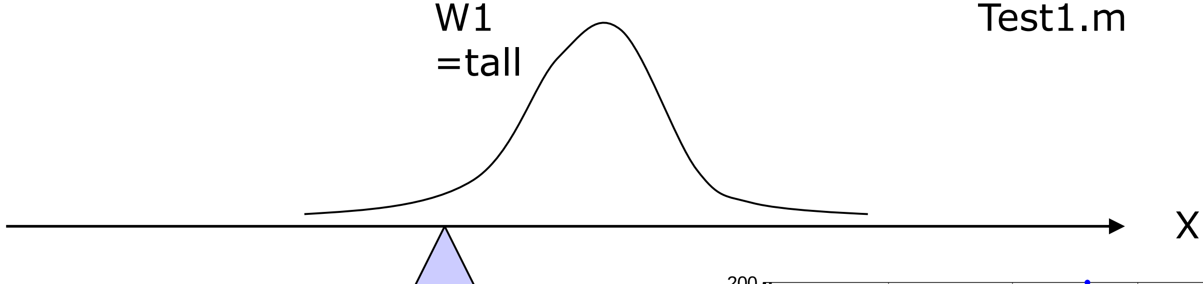
$$x \in X$$

Who will determine it?

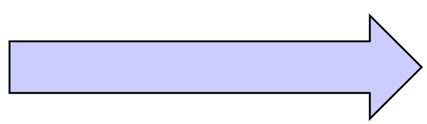
Classification: Bayesian Classifier

- Random variable, x : probability of event occurrence.
- When x is given, is x involved in class w1 or w2?

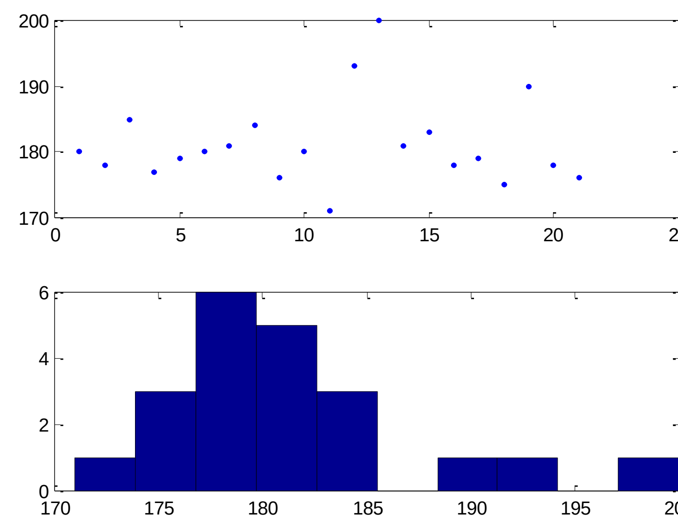
Test1.m



$$x \in X$$

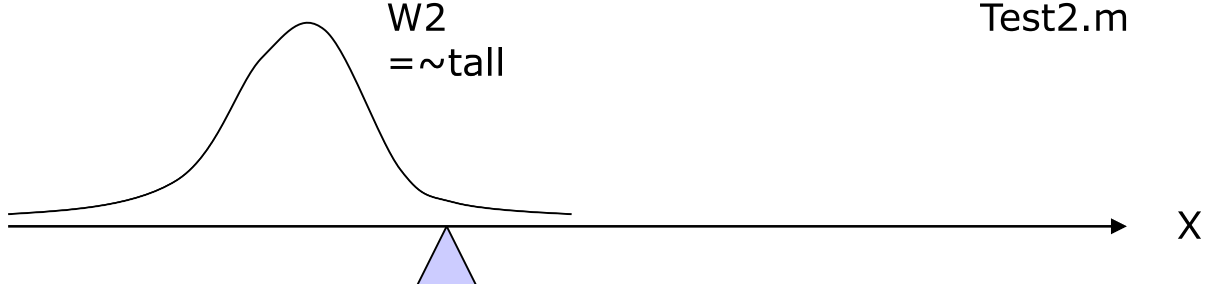


H=180,   I think tall.



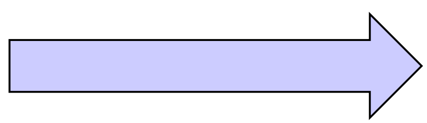
Classification: Bayesian Classifier

- Random variable, x : probability of event occurrence.
- When x is given, is x involved in class w1 or w2?

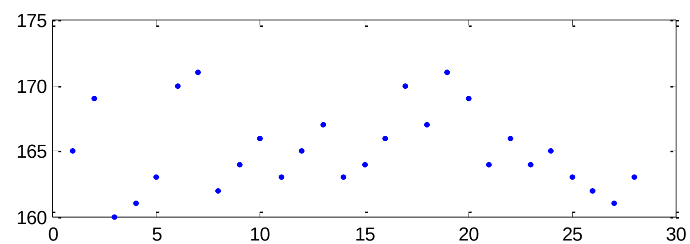


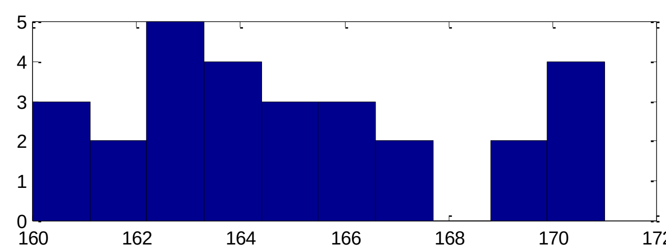
Test2.m

$$x \in X$$

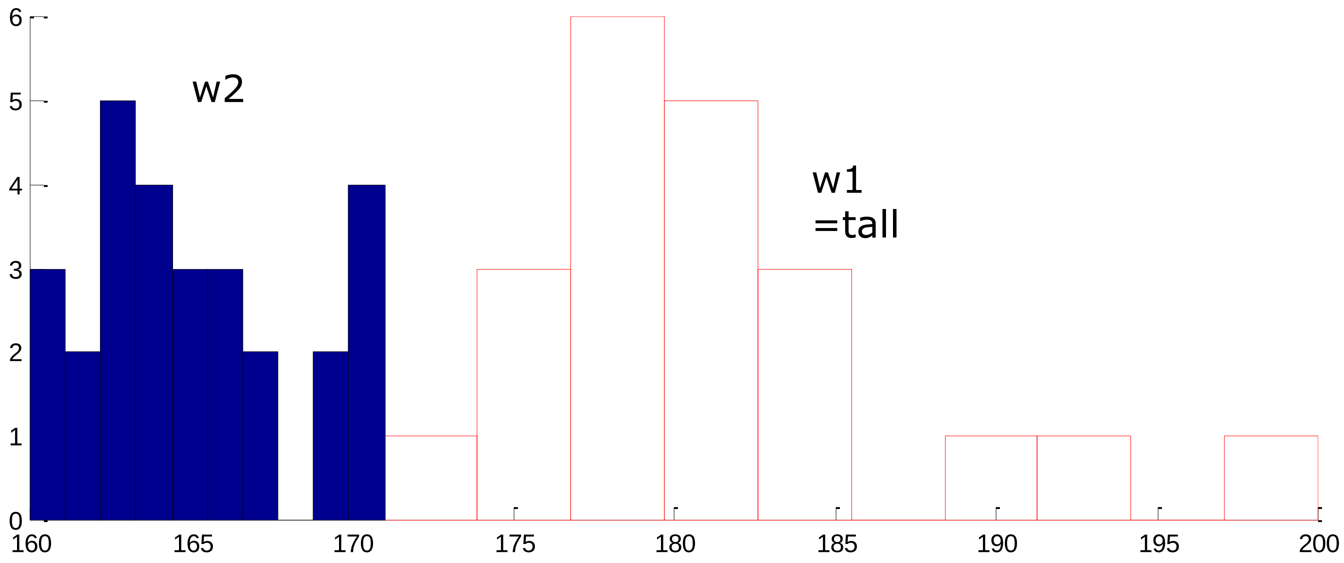


H=160,   I think ~tall





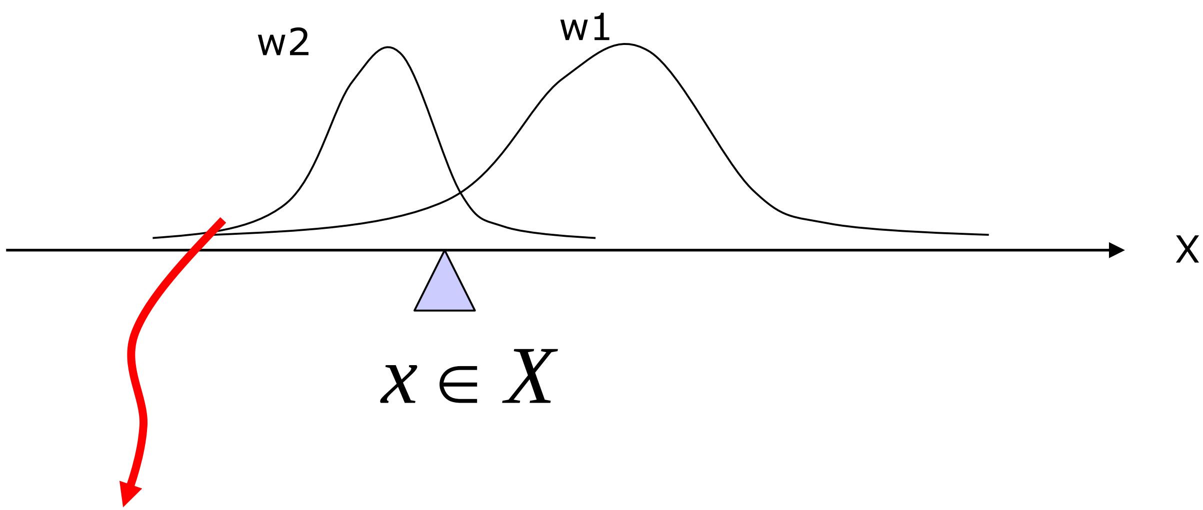
Samples from Surveys.



- Assume that samples have Gaussian distribution.
- (m1,s1) = (181.143, 6.54)
- (m2,s2) = (165.14, 3.12)

Classification: Bayesian Classifier

- Random variable, x : probability of event occurrence.
- When x is given, is x involved in class w1 or w2?



$$x \in X$$

$$\Pr(x) = ?$$

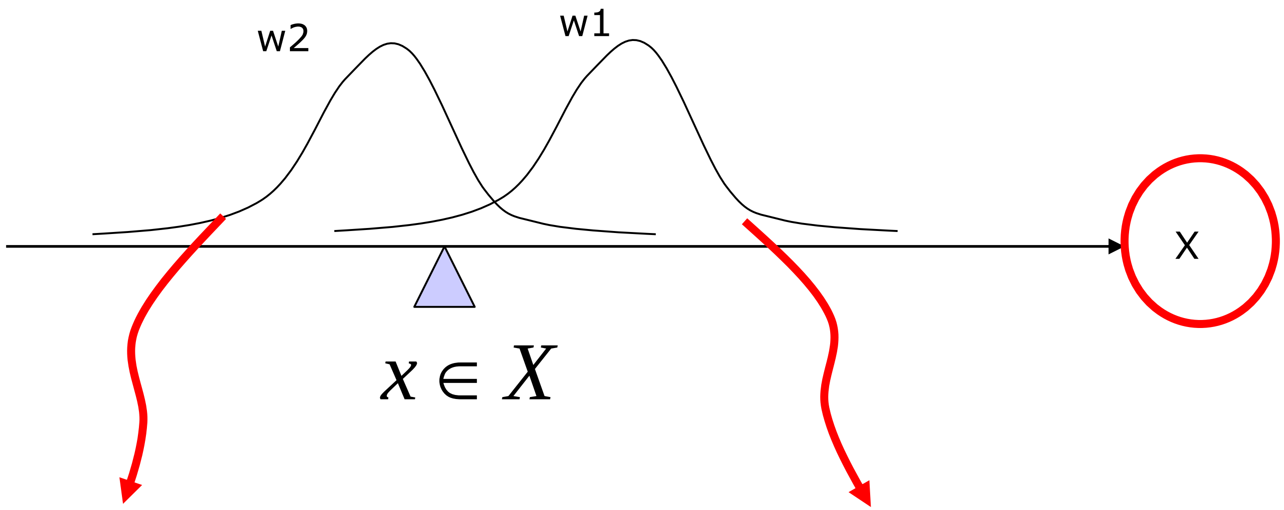
$$\Pr(w_2 \mid x) = ?$$

$$\Pr(w_2) = ?$$

$$\Pr(x \mid w_2)$$

Classification: Bayesian Classifier

- Random variable, x : probability of event occurrence.
- When x is given, is x involved in class w1 or w2?



$$x \in X$$

$$P(x \mid w_2)$$

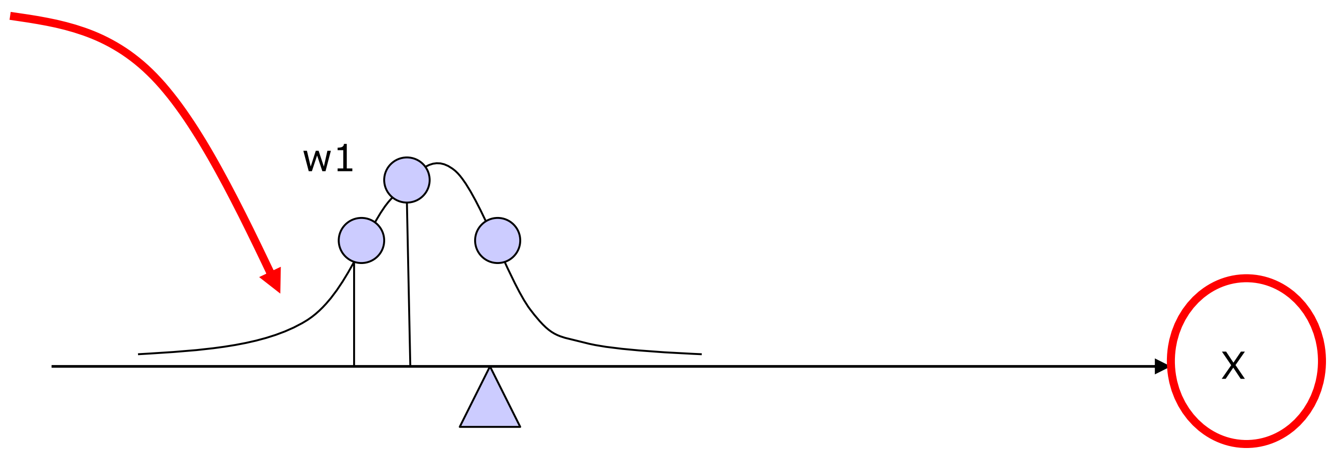
In w2 group, samples xs are gathered.

$$P(x \mid w_1)$$

In w1 group, samples xs are gathered.

Samples.

```
x=[ 180
   178
   185
   177
   179
   180
   181
   184
   176
   180
   171
   193
   200
   181
   183
   178
   179
   175
   190
   178
   176];
```



$$x \in X$$

$$\therefore \Pr(x \mid w_1)$$

For Bayesian Classifier, p(x|w) and p(w) are required.

- How to find P(w)?

n(w1) =21

```
x=[ 180
    178
    185
    177
    179
    180
    181
    184
    176
    180
    171
    193
    200
    181
    183
    178
    179
    175
    190
    178
    176];
```

n(w2) =28

```
x2=[165
    169
    160
    161
    163
    170
    171
    162
    164
    166
    163
    165
    167
    163
    164
    166
    170
    167
    171
    169
    164
    166
    164
    165
    163
    162
    161
    163];
```

P(w1) =21/(21+28)

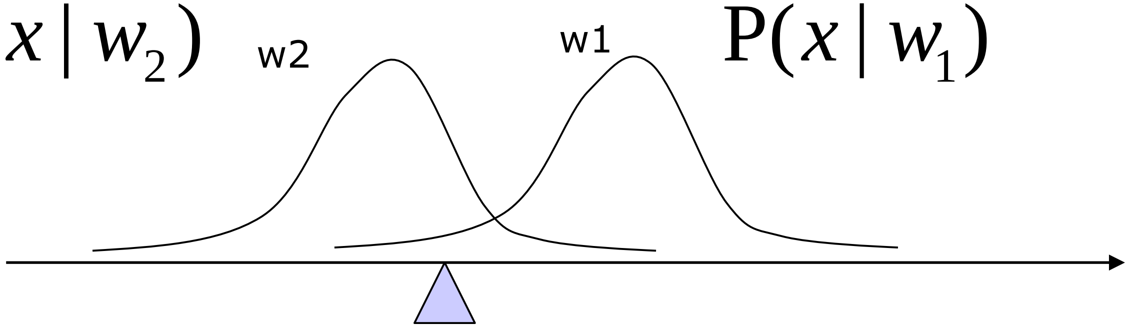
P(w2) =28/(21+28)

P(w1)+P(w2) =1

Back to Bayesian Probability

- x is given, is it w1 or w2?

$$\mathrm{P}(x \mid w_2) \qquad\qquad \mathrm{P}(x \mid w_1)$$



$$x \in X$$

$$\mathrm{P}(A \mid B) = \frac{P(\mathrm{A} \wedge \mathrm{B})}{P(B)}$$

$$P(A \mid B)P(B) = P(A \wedge B) = P(B \mid A)P(A)$$

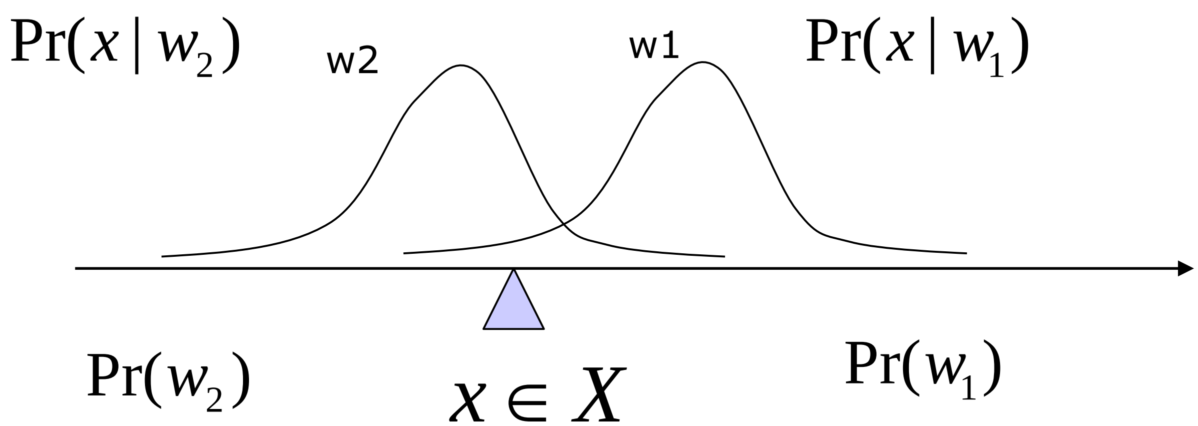
$$\therefore \Pr(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

A→ w1, B→x
A→ w2, B→x

$$\therefore \mathrm{P}(w_1 \mid x) = \frac{P(x \mid w_1)\, p(w_1)}{P(x)}$$

$$\therefore \mathrm{P}(w_2 \mid x) = \frac{P(x \mid w_2)\, p(w_2)}{P(x)}$$

Definition
Bayesian Classifier



$$\Pr(w_1 \mid x) = \frac{P(x \mid w_1)\, p(w_1)}{P(x)} > \Pr(w_2 \mid x) = \frac{P(x \mid w_2)\, p(w_2)}{P(x)}$$

$$\textit{then}, \qquad x \in w_1$$

$$\textit{otherwise}, \qquad x \in w_2$$

Finally, p(x)=?

$$\mathrm{P}(w_1 \mid x) = \frac{P(x \mid w_1)\, p(w_1)}{P(x)}$$

$$\rightarrow 1 = \mathrm{P}(w_1) + \mathrm{P}(w_2)$$

$$\therefore \mathrm{P}(x) = \mathrm{P}\left(x \mid w_1\right)\mathrm{P}(w_1) + \mathrm{P}(x \mid w_2)\,\mathrm{P}(w_2)$$

- Finally,

$$\mathrm{P}(w_1 \mid x) = \frac{P(x \mid w_1)\, p(w_1)}{P(x \mid w_1)\, p(w_1) + P(x \mid w_2)\, p(w_2)}$$

Posterior Probability

Posterior Probability in General

$$P(w_i \mid x) = \frac{P(x \mid w_i)\, p(w_i)}{P(x) = P(x \mid w_1)\, p(w_1) + P(x \mid w_2)\, p(w_2)}$$

$$\rightarrow \mathrm{P}(w_i \mid x) = \frac{P(x \mid w_i) P(w_i)}{\sum_k P(x \mid w_k) P(w_k)}$$

$$when \quad \sum_k P(w_k) = 1$$

$$Warning!...\sum_k P(x \mid w_k) P(w_k) \neq 1$$

Engineering Notation

$$\mathrm{P}(\mathrm{w} \mid \mathrm{x}) = \frac{P(\mathrm{x} \mid \mathrm{w}) P(\mathrm{w})}{P(\mathrm{x})}$$

$$Posterior = \frac{likelihood \times prior}{Evidence}$$

In engineering, likelihood is one of the popular solution.

What is the difference between Likelihood and Posterior probability?

$$P(\mathrm{w}\,|\,\mathrm{x}) = \frac{P(\mathrm{x}\,|\,\mathrm{w})P(\mathrm{w})}{P(\mathrm{x})}$$

- likelihood-based classifier

$$P(x\,|\,w_1) > P(x\,|\,w_2)$$

$$then\ x \in w_1$$

대충 키 큰사람은
평균이 181,
작은 사람은 165
이니,
X= 175는 키가 큰쪽에
확률에 가깝다?

- Posterior probability-based classifier

$$P(w_i\,|\,x) = \frac{P(x\,|\,w_i)P(w_i)}{P(x\,|\,w_1)P(w_1) + P(x\,|\,w_2)P(w_2)}$$

$$P(w_1\,|\,x) > P(w_2\,|\,x)$$

$$then,\quad x \in w_1$$

X=175인 경우,
키가 클 확률은 얼마
작을 확률은 얼마이므로
키다 크다 또는 작다..

Example: Test4.m

$$x \sim N(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

- N(m,s) → normpdf in Matlab
- x =175 → N(m,s,x=175)

$$p(x=175 \mid w_1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left[-\frac{(175-\mu_1)^2}{2\sigma_1^{\,2}} \right]$$

- Likelihood prob. classifier

$$P(x \mid w_1) = pxw1 = 0.0392$$

$$P(x \mid w_2) = pxw2 = 0.0009$$

- Posterior prob. classifier

$$P(w_1 \mid x) = pw1x = 0.971$$

$$p(w_2 \mid x) = pw2x = 0.029$$

Theoretical Interest

- We can think error.

$$P(error \mid x) = \begin{cases} p(w_1 \mid x) & \text{if we decide } w_2 \\ p(w_2 \mid x) & \text{if we decide } w_1 \end{cases}$$

$$P(e) = P(e \mid x) P(x)$$

- Thus, from posterior classifier, we define p(e)

$$P(w_1 \mid x) > P(w_2 \mid x) \rightarrow$$

$$Bayesian\ classifier: \ \arg\max(p(w_i \mid x))$$

$$Bayesian\ Error: \ \arg\min(p(w_i \mid x))$$

Bayesian Error is,

- Very small.
- In many cases, Bayesian classifier is better than you.

- Most classifiers are compared with Bayesian error.
- If you have success of designing new classifier,

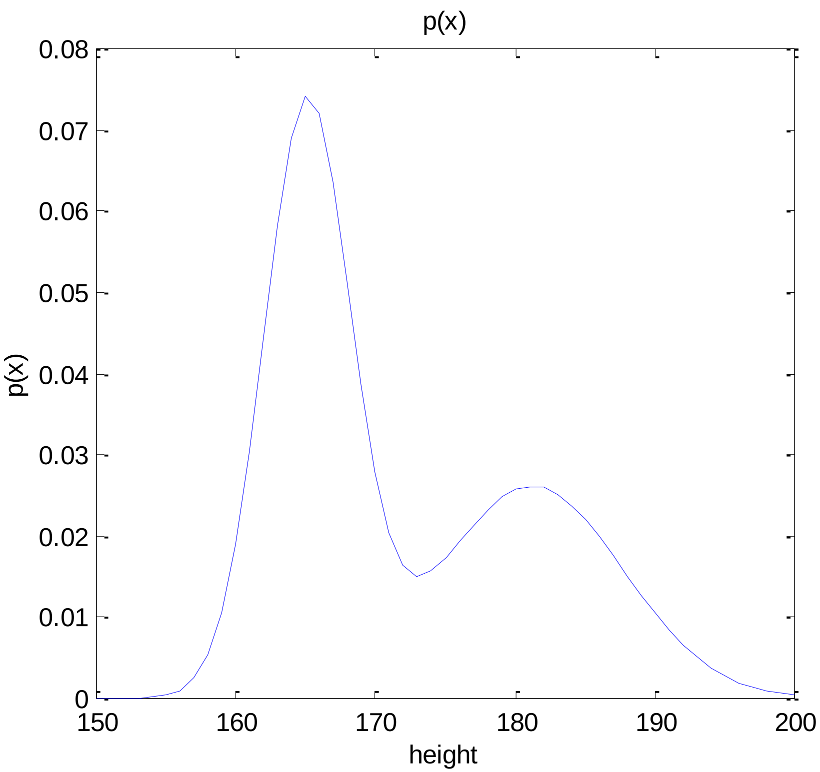
in general, its performance is probably rather better than Bayesian classifier.

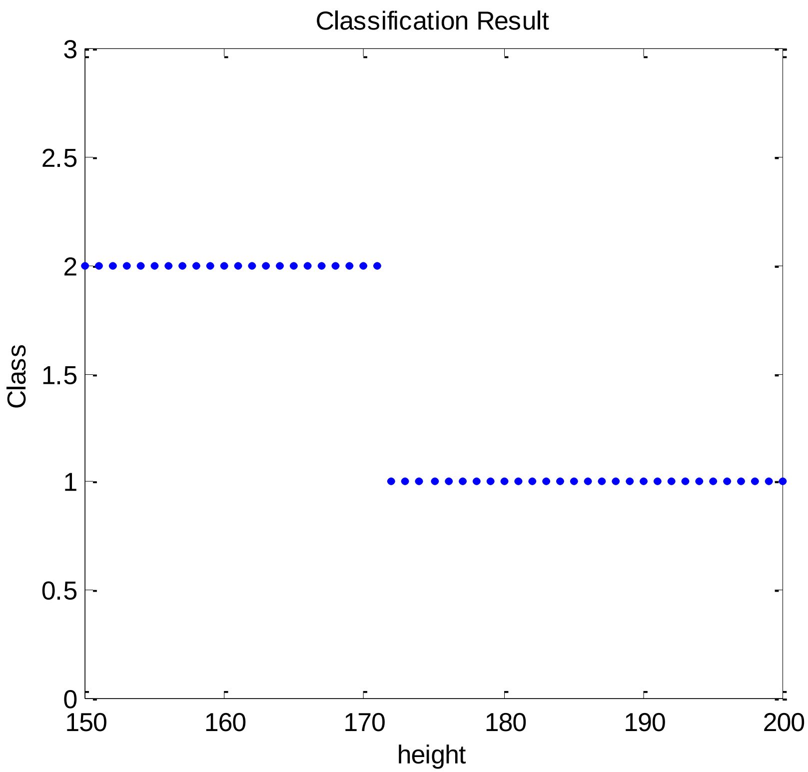
- Mathematically, Bayesian classifier is VERY STRONG.
  - Question: Why Deep Learning is so good?
  - Because, DL has the function of finding GOOD Feature.

Example. Test 5.m
Plot everything

- P(x) $\therefore \mathrm{P}(x) = \mathrm{P}\left(x \mid w_1\right)\mathrm{P}(w_1) + \mathrm{P}(x \mid w_2)\,\mathrm{P}(w_2)$

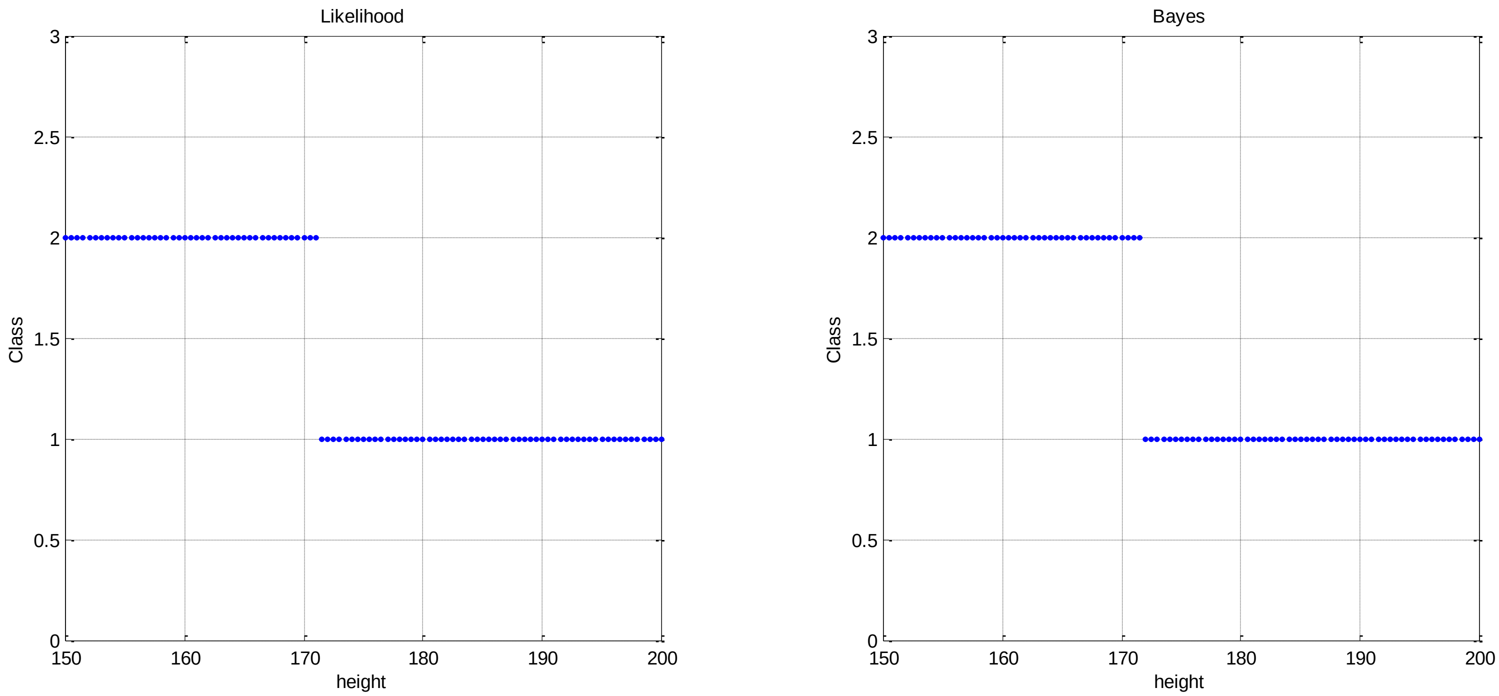
$$\textit{Bayesian classifier}: \quad \arg\max(p(w_i \mid x))$$





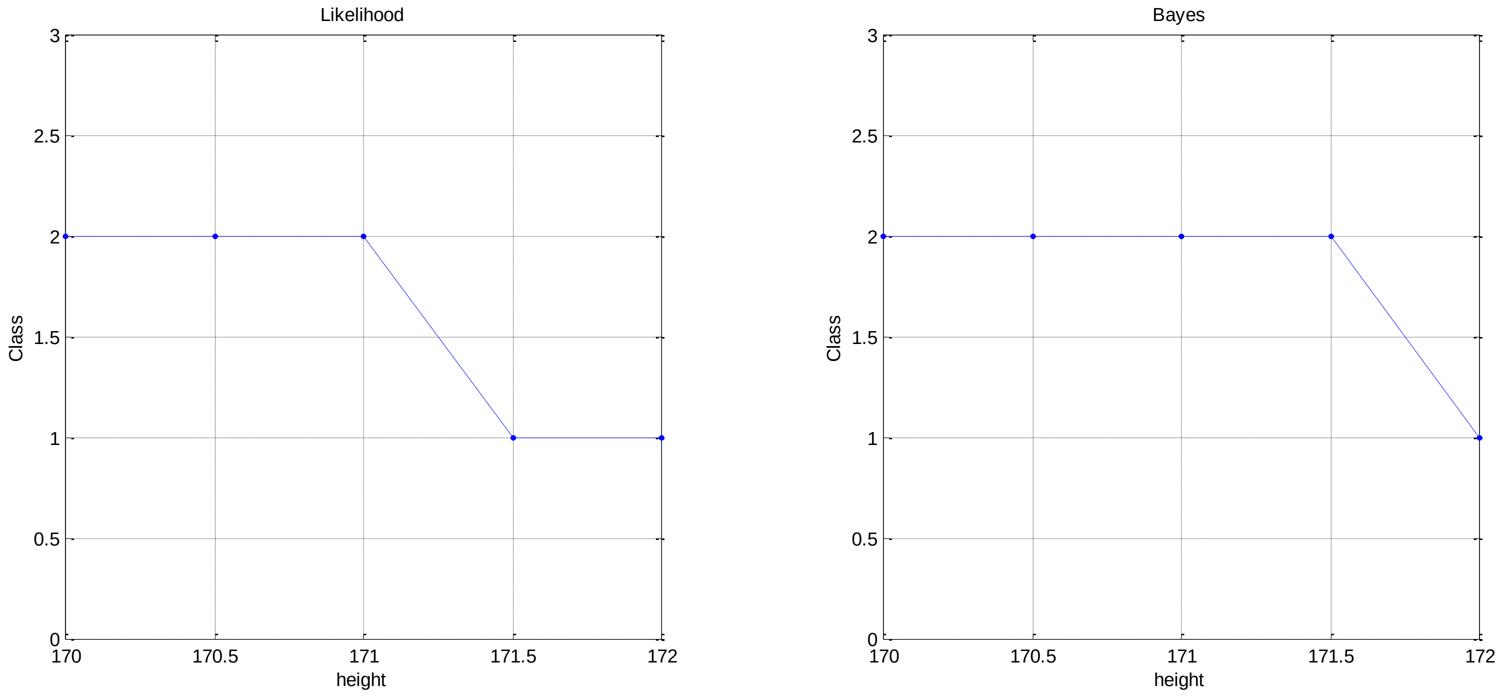
Likelihood Vs. Posterior

$$i = \arg\max(P(x \mid w_i)) \quad \text{vs.} \quad i = \arg\max(P(w_i \mid x))$$



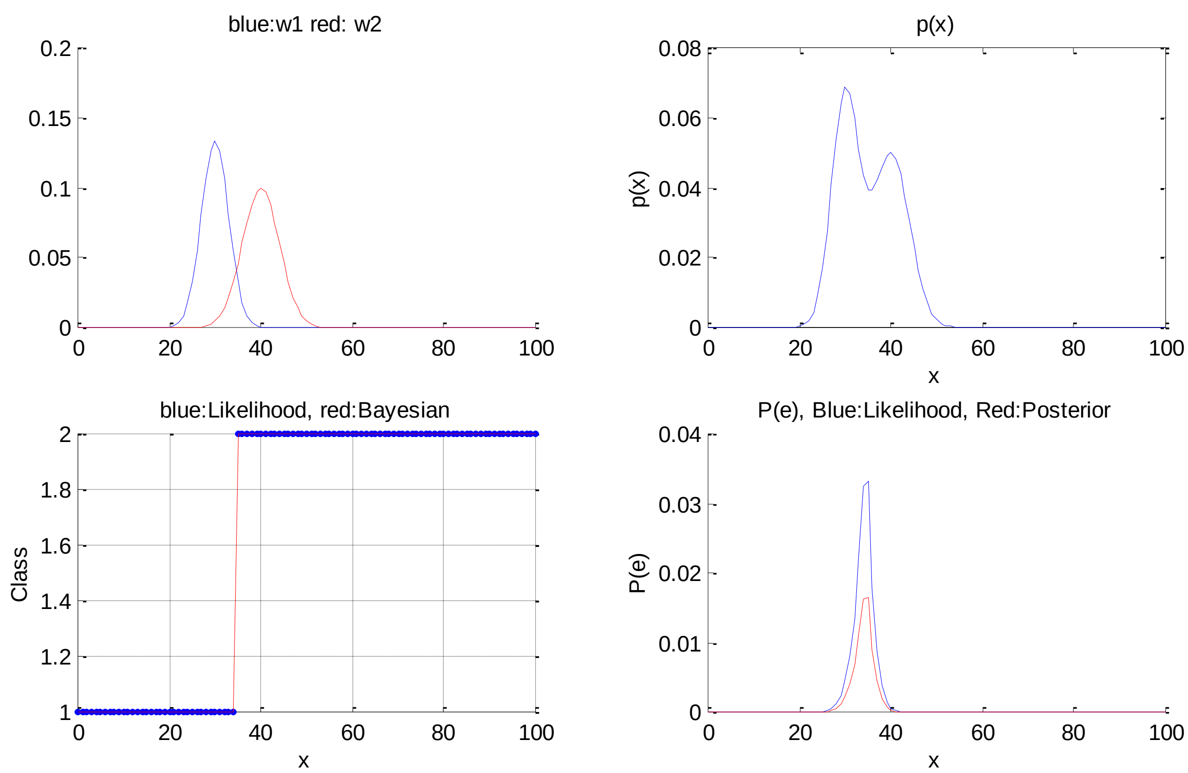
Equal?

Overlapped Area



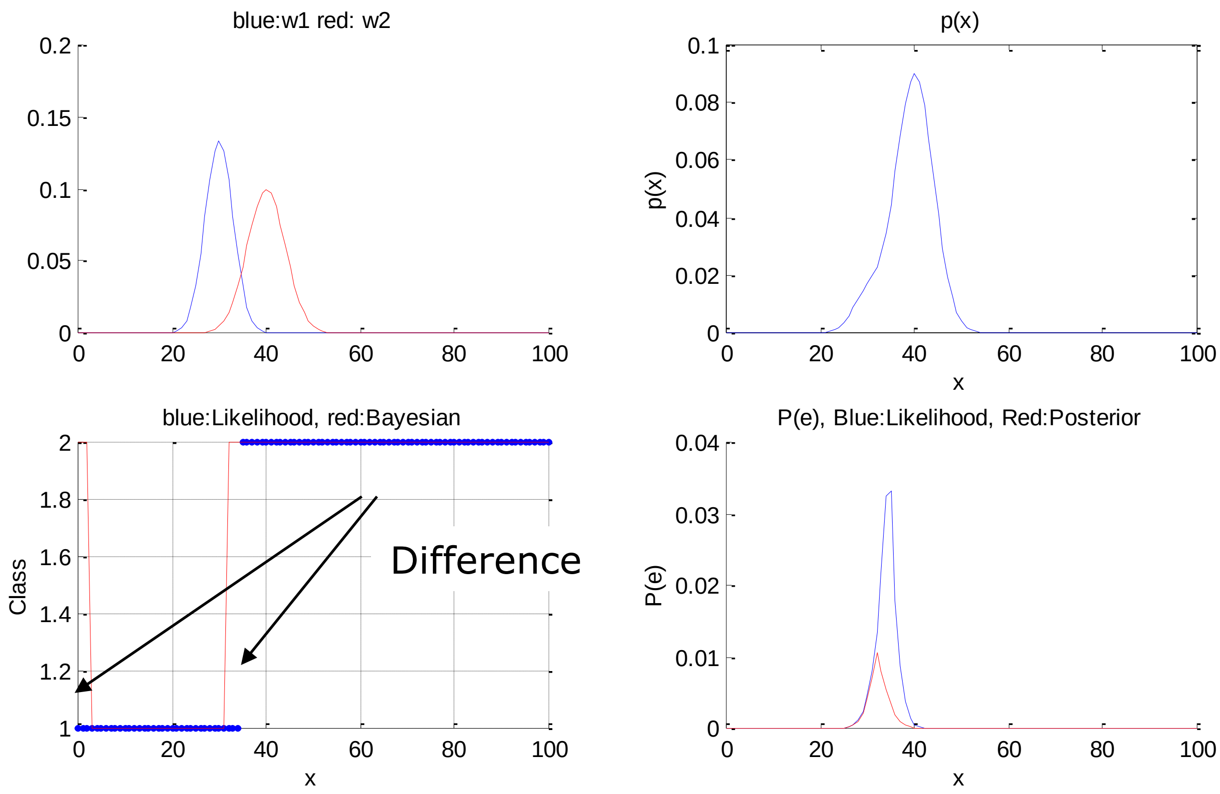
The Most Important Factor of Classifier. → Minimize Error on Overlapped data

- New Data in Test 6.m
- w1~N(30,2) w2~N(40,4)   p(w1)=p(w2)=0.5



Problems of Likelihood. When $P(w_1) \neq P(w_2)$

- Likelihood CANNOT be used for $P(w_1) \neq P(w_2)$
- Ex) P(w1)=0.1 P(w2)=0.9



Example of Bayesian Classifier
: Sensor for Something

- Example of PSD (distance sensor)
- If (distance>0.8) then “human exists” else “nothing”.

Returns “distance”

Who will choose threshold?

- 1. Adhoc
  - Well, 0.8 could be the possible value.
  - You will go to Jail… T_T..
- 2. Likelihood
  - After 100 samples,

$$P(x \mid w_{human}) > P(x \mid w_{nothing}) \rightarrow \text{Human}$$

  - When samples are not balanced… it also fails.
- 3. Bayesian
  - After 100 samples

$$P(w_{human} \mid \mathrm{x}) > P(w_{nothing} \mid \mathrm{x}) \rightarrow \text{Human}$$

  - You did your best except for Deep Learning..^_^…

Specification of Bayesian Classification

- Bayesian classification
  - It requires a lot of Samples
  - Everything are designed with Probabilistic Distribution
  - **Therefore, Modeling-based Method(Parametric Method)**
- When class is added in online environment, it is useless.
  - But, most classifications are useless, too.
- When new samples cannot be used.
  - After sampling, Bayesian classifier is calculated.
- Any method in which New samples are updated,
  - Non parametric method(usually, Kernel based method)

Classification and Features

- x is a random variable.
- But, x is also called as a feature vector.

- In a given problem, **you should find a good feature**.
- Grade, creativity, moral attention could be features for recruiting students.

- PSD distance is not enough. Movements could be OK.

- **Without GOOD features, classifier cannot work.**