

機械学習特論

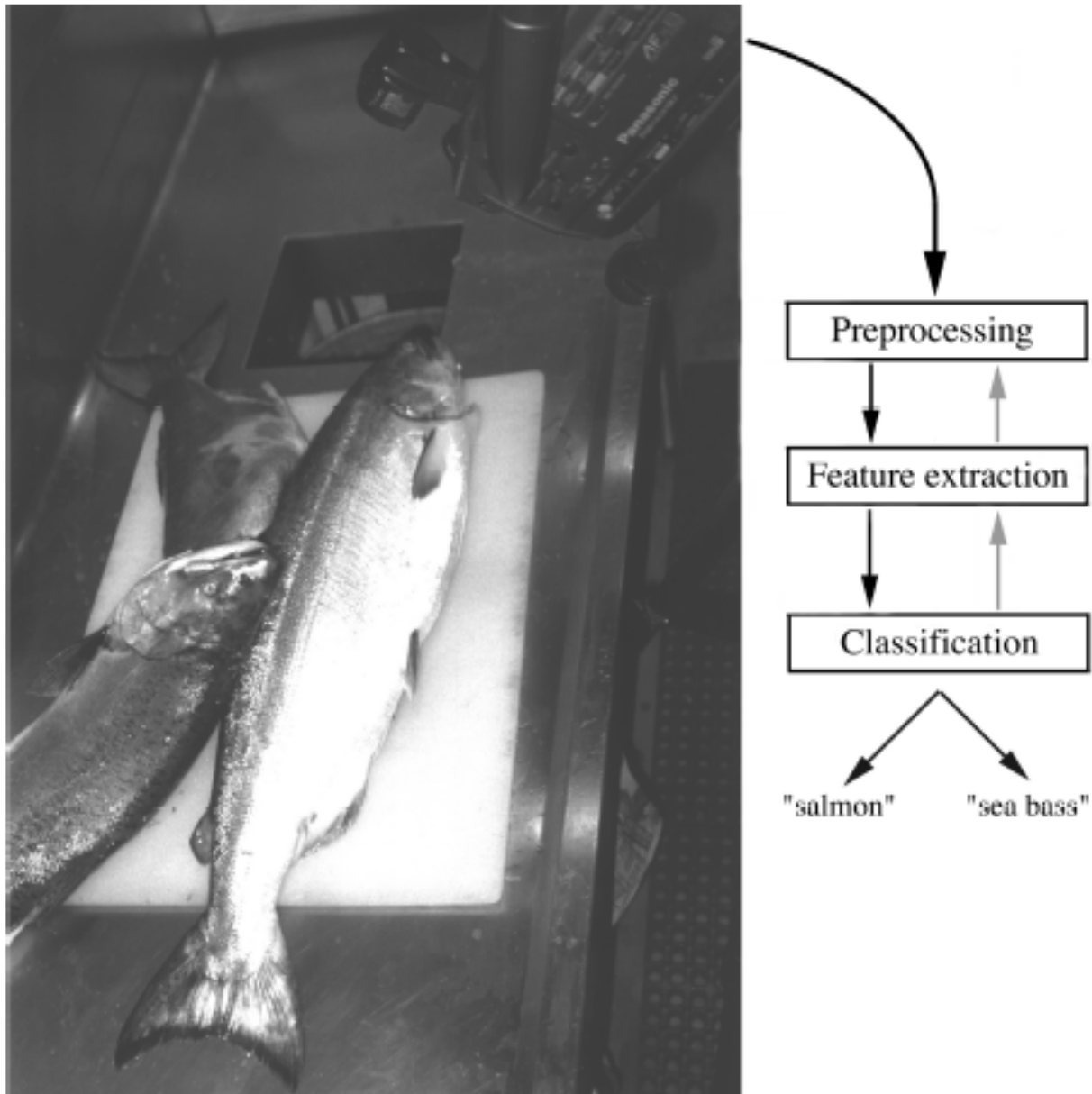
~理論とアルゴリズム~

(Basics in probabilities)

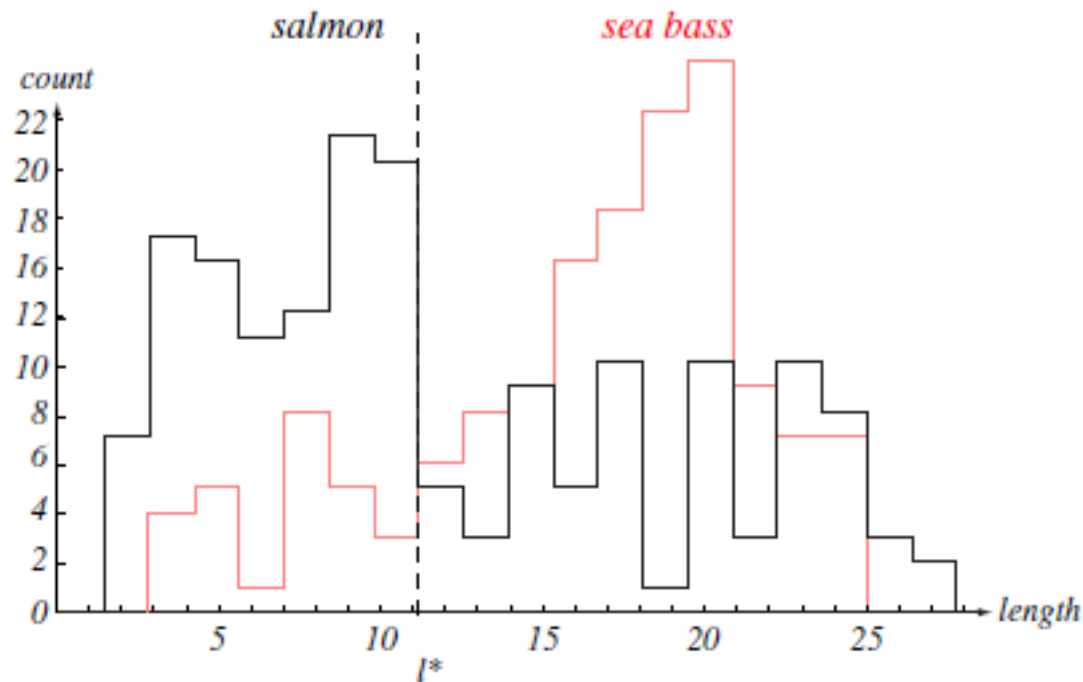
講師：西郷浩人

Goal of pattern recognition:
Parametric (rule-based) approach

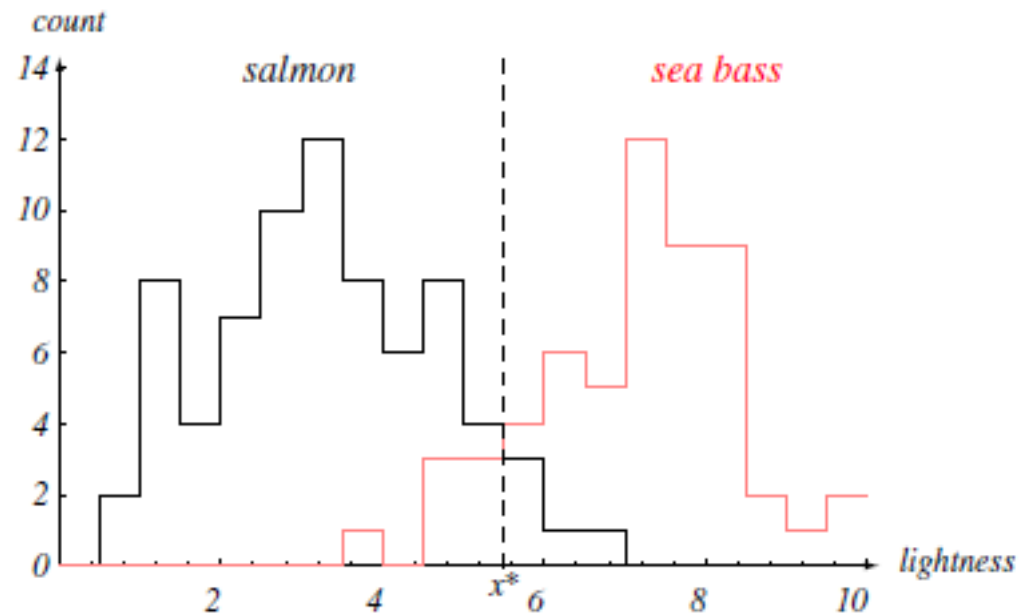
Process of Learning



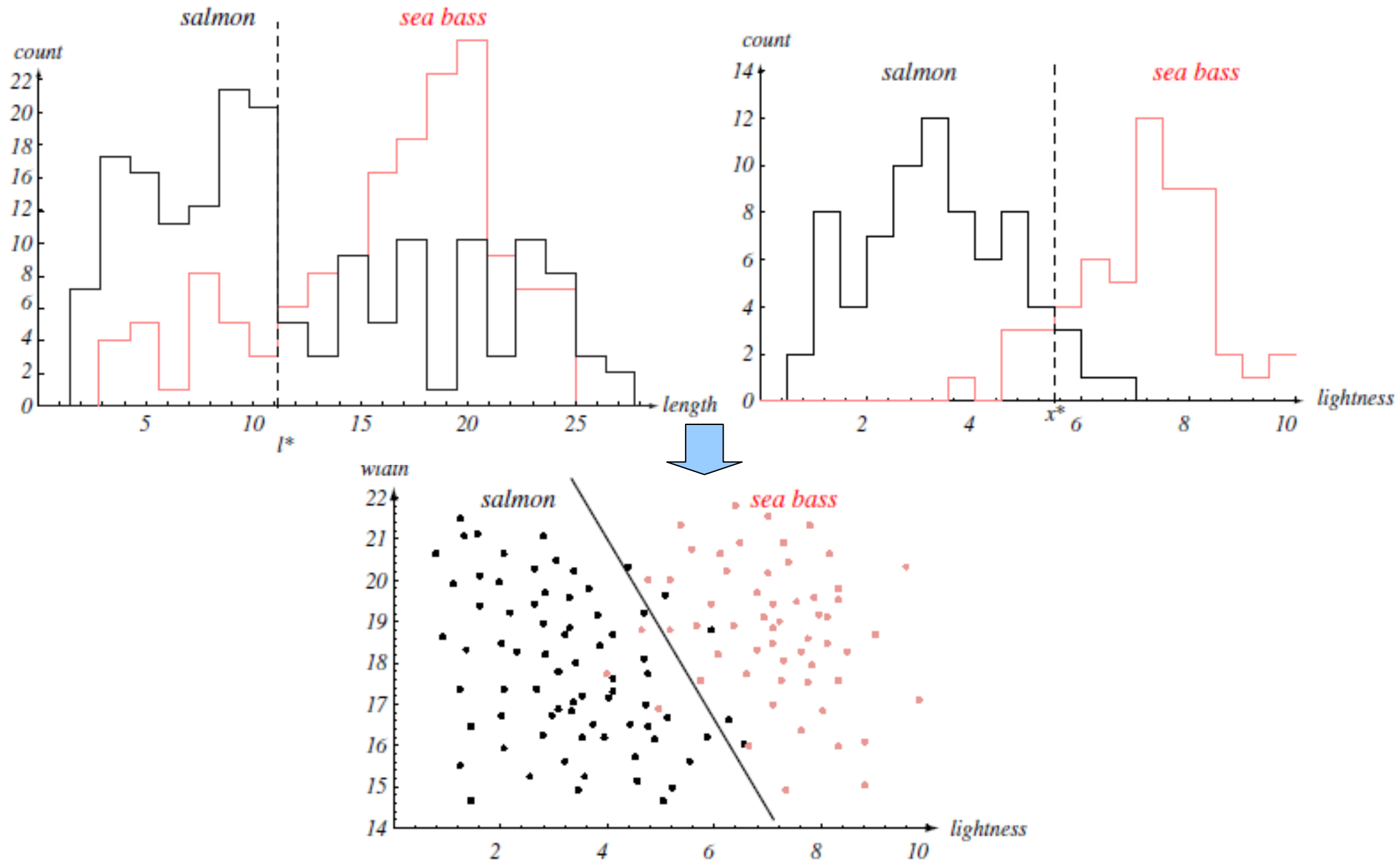
Classification by the fish length



Classification by the brightness



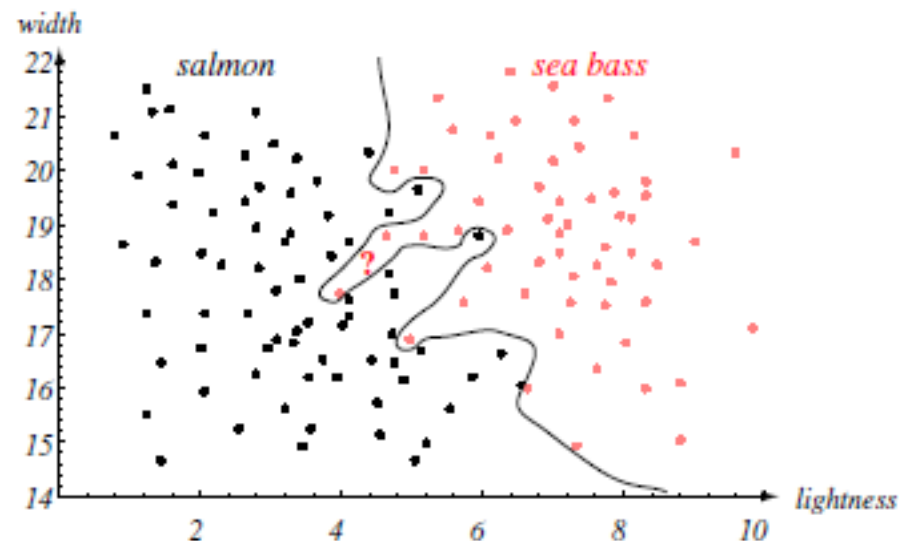
Classification by the combined rule.



Can we just increase the number of
rules (features) ?

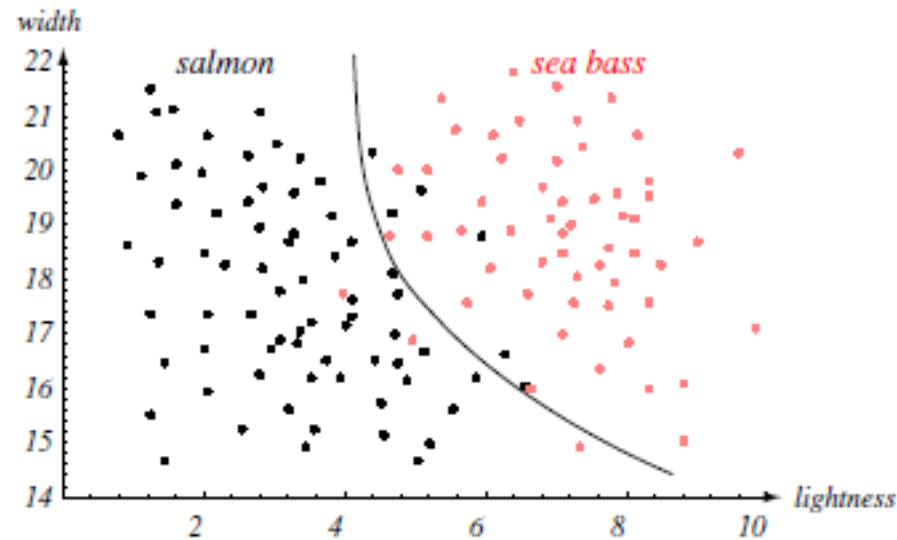
No

- Too many rules complicates the classification boundary (below), which can be prone to a newly available data point. (overfitting)



Good rule

- should be simple and robust to a newly available data point.



Learning rules

Consider extracting rules for discriminating salmon and sea bass, such as size, color, etc) out of limited amount of data.

Our data can be represented in the following way.
x represents rules (features), y represents salmon(1) or sea bass(0).

$$\{\{\mathbf{x}_1, y_1\}, \{\mathbf{x}_2, y_2\}, \dots, \{\mathbf{x}_n, y_n\}\}$$

Learning rules continued.

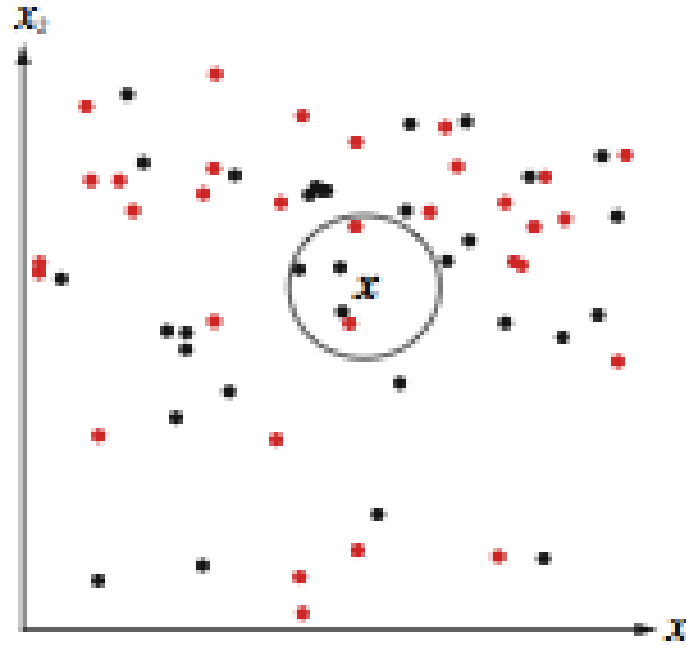
We will see different formulation and motivation for this objective. Examples covered in this lecture series are;

- 1) LDA
- 2) Logistic Regression
- 3) Neural Networks
- 4) SVM

We will see them later.

Non-parametric (sample-based)
approach

k-nearest neighbor



- Procedure
 - Choose k training data points closest to the test point, then perform majority vote.
 - k should be an odd number such as 1, 3, 5, ... etc.

Review on basic probabilities

random variable (確率変数)

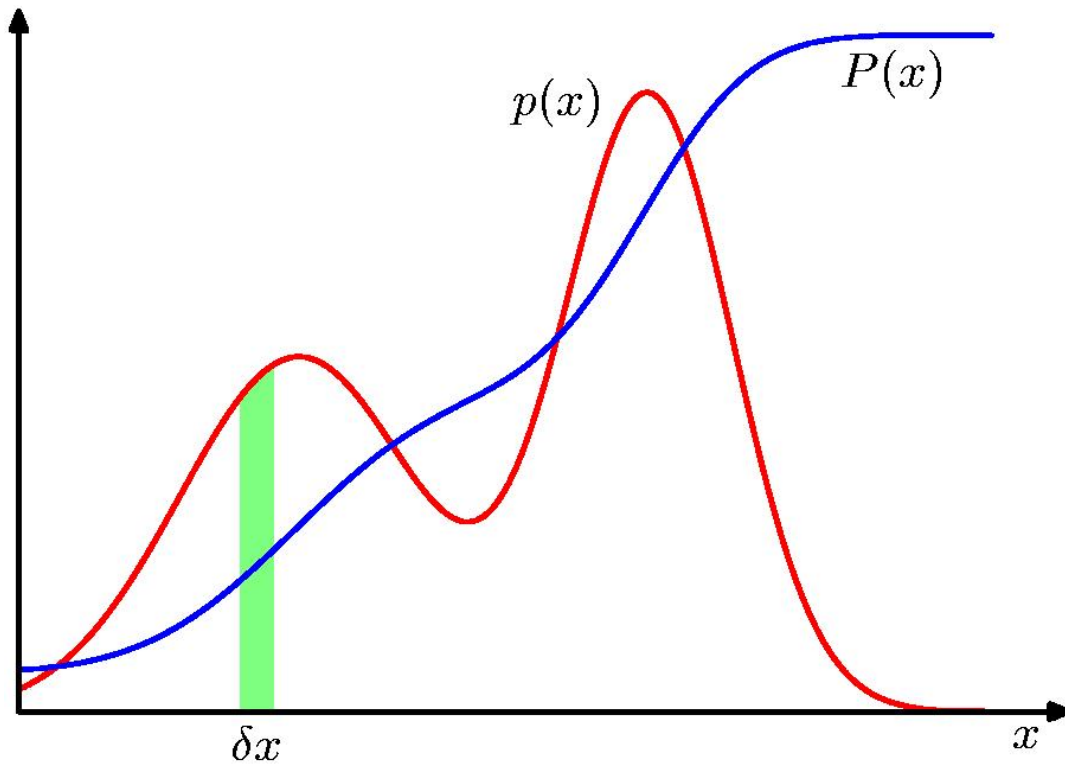
□
Response variable y is a discrete random variable, and satisfies the following condition.

$$\sum_{i=1}^k p(y_i) = 1, \quad p(y_i) \geq 0$$

Explanatory variable (feature) x is generally a continuous variable, and satisfies the following condition.

$$\int p(x) dx = 1, \quad p(x) \geq 0$$

Probability function $P(x)$ and probability density function $p(x)$



$$\int_{-\infty}^{\infty} p(x) dx = 1, \quad p(x) \geq 0$$

$$P(Z) = \int_{-\infty}^Z p(x) dx$$

$$P(x \in a, b) = \int_a^b p(x) dx$$

Joint and conditional probability

Joint probability (同時確率) between x and y

$$p(x, y)$$

Conditional probability(条件付き確率) of y given x

$$p(y|x)$$

The Rules of Probability

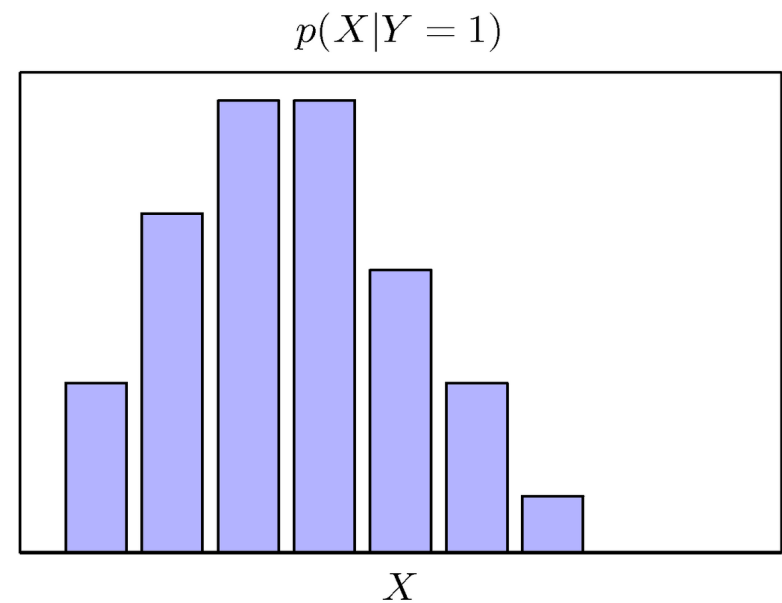
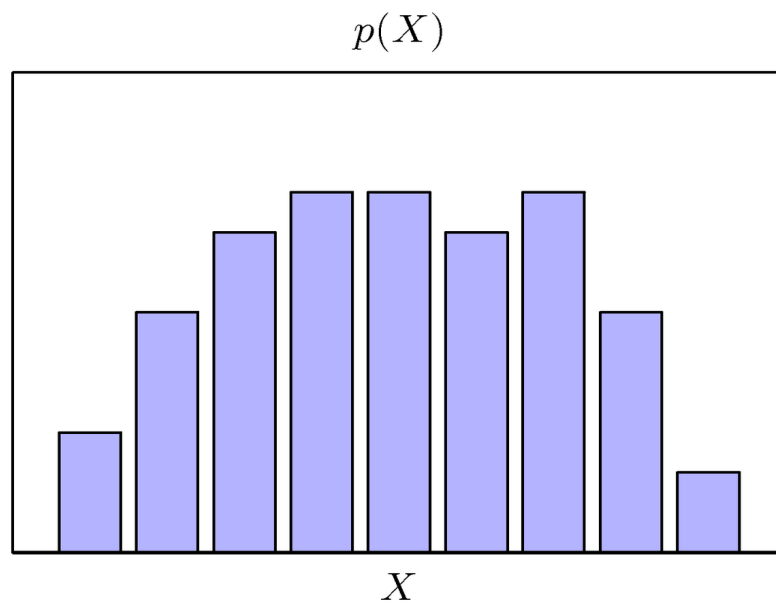
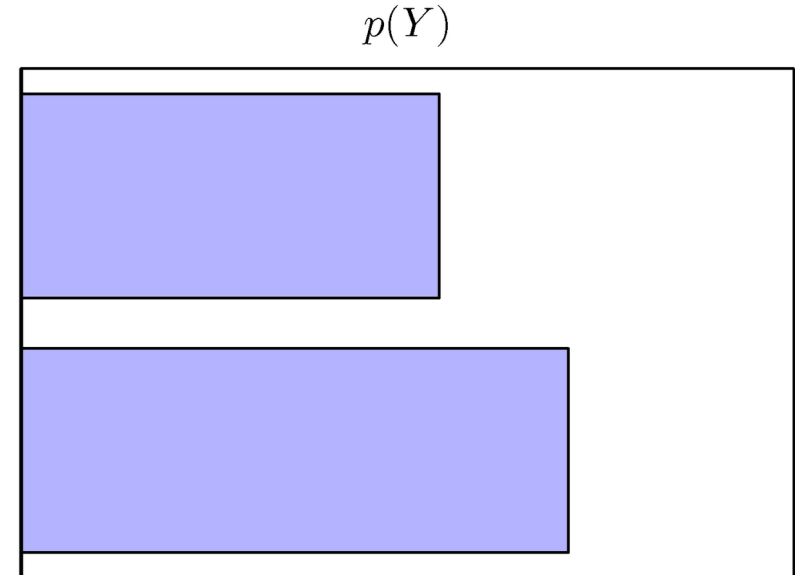
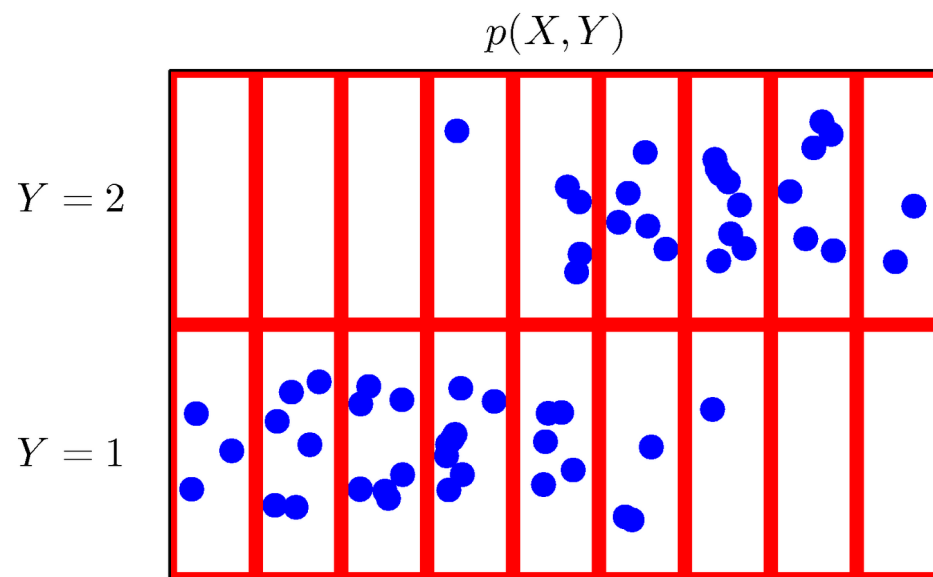
Sum Rule

$$p(X) = \sum_Y p(X, Y)$$

Product Rule

$$p(X, Y) = p(Y|X) p(X)$$

Example: marginal and conditional probability



Bayes' theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$$

$p(Y|X)$: posterior probability (事後確率)

$p(X|Y)$: likelihood (尤度)

$p(Y)$: prior probability (事前確率)

$p(X)$: データの確率

The denominator is often skipped, and described as

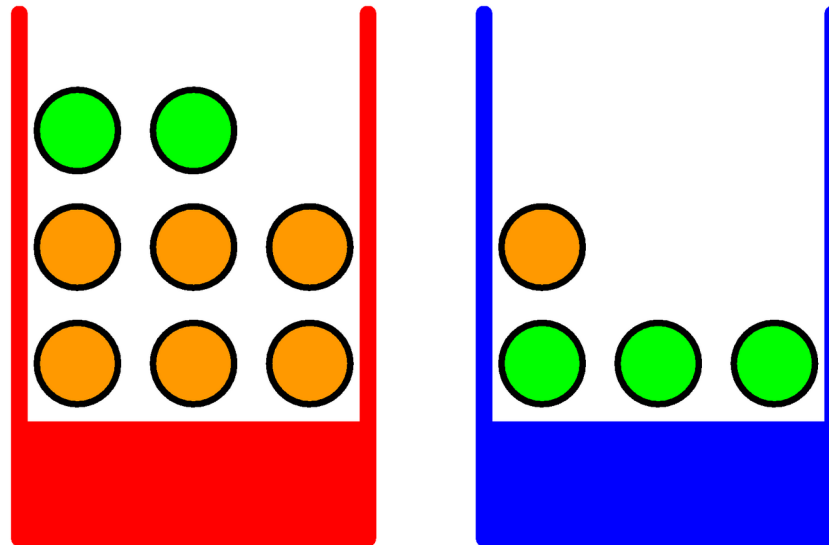
$$p(Y|X) \approx p(X|Y)p(Y)$$

Example Bayes' theorem

Red box: 6 oranges and 2 apples

Blue box: 1 orange and 3 apples

Probability to choose red box: 40%, blue box: 60%



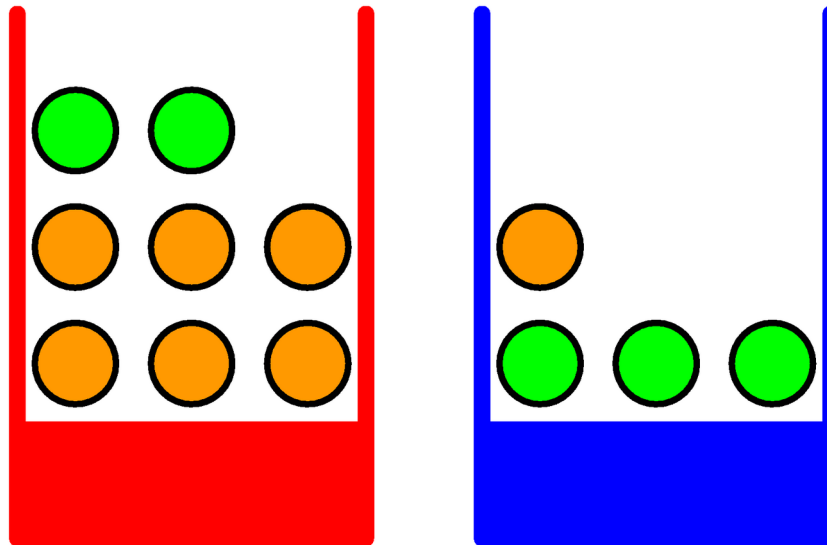
Example: Bayes' theorem

$$p(B=r)=4/10 \quad p(B=b)=6/10$$

$$p(F=a|B=r)=1/4 \quad p(F=o|B=r)=3/4$$

$$p(F=a|B=b)=3/4 \quad p(F=o|B=b)=1/4$$

What's the probability of $B=r$ when $F=o$?



What's the probability of $B=r$ when $F=o$?

$$p(B=r)=4/10 \quad p(B=b)=6/10$$

$$p(F=a|B=r)=1/4 \quad p(F=o|B=r)=3/4$$

$$p(F=a|B=b)=3/4 \quad p(F=o|B=b)=1/4$$

$$\begin{aligned} & p(B=r|F=o) \\ &= \frac{p(F=o|B=r) p(B=r)}{p(F=o)} \\ &= \frac{p(F=o|B=r) p(B=r)}{p(F=o|B=r) p(B=r) + p(F=o|B=b) p(B=b)} \\ &= \frac{3/4 \times 4/10}{3/4 \times 4/10 + 1/4 \times 6/10} = \frac{2}{3} \end{aligned}$$

Other statistics

- 平均(mean) $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

- 分散(variance) $var(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

- 標準偏差 $sd(\mathbf{x}) = \sqrt{var(\mathbf{x})}$

(standard deviation)

- 共分散 $cov(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

(covariance)

- 相関 $cor(\mathbf{x}, \mathbf{y}) = \frac{cov(\mathbf{x}, \mathbf{y})}{sd(\mathbf{x})sd(\mathbf{y})} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$

(correlation)

- 期待値 $E(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i p(x_i)$

(expectation)

共分散(covariance)と相関(correlation)

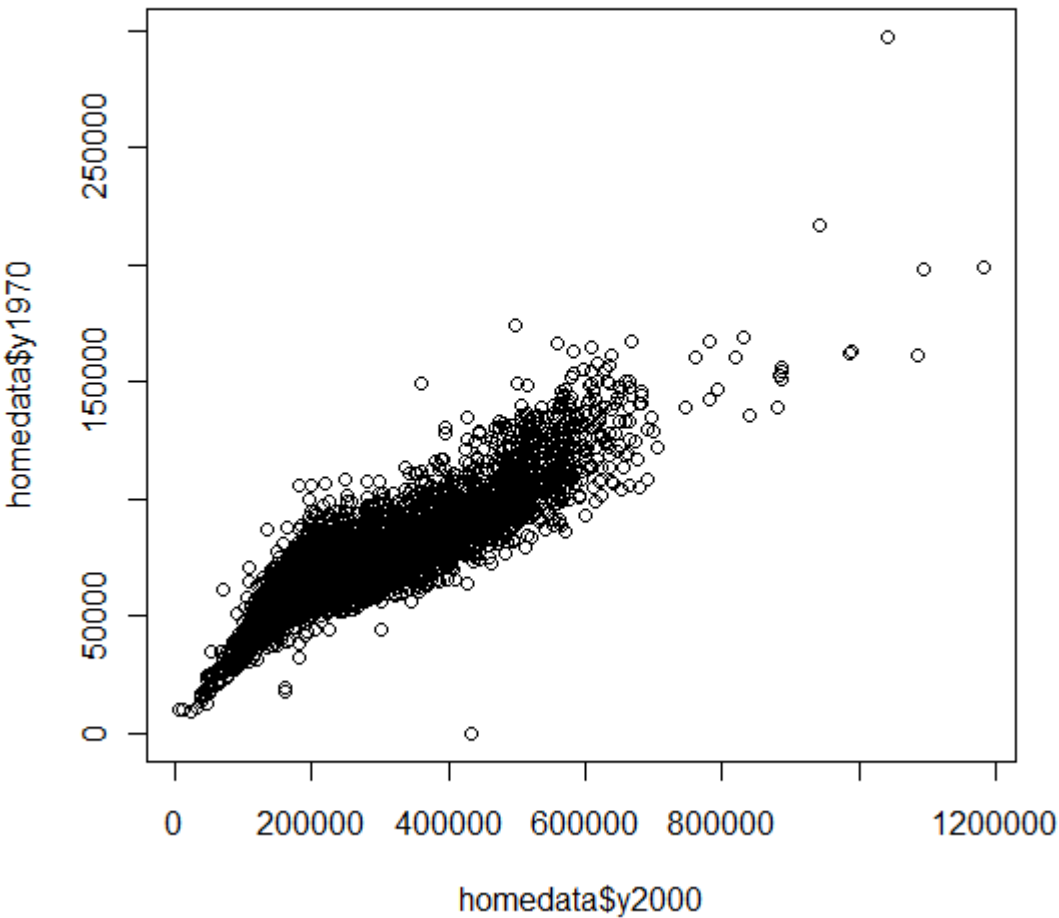
correlation(相関係数) is a normalized covariance

$$-1 \leq \text{cor} \leq 1, -\infty < \text{cov} < \infty$$

House price 2000 vs 1970

Cov = 2610880277

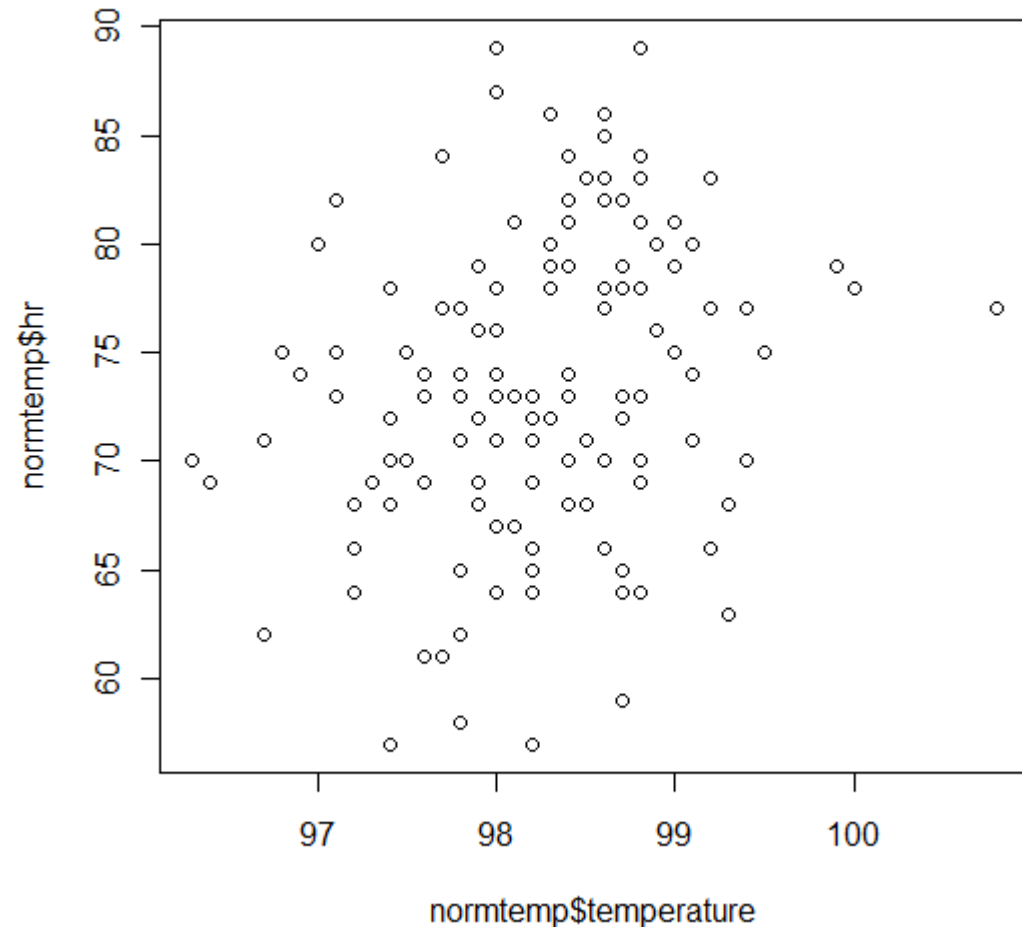
Cor = 0.8926155 (high)



temp vs heart rate

Cov = 1.313381

Cor = 0.2536564 (low)



Ex. 1

Suppose that we have three coloured boxes r (red), b (blue) and g (green). Box r contains 3 apples, 4 oranges and 3 limes, box b contains 1 apple, 1 orange and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r)=0.2$, $p(b)=0.2$, $p(g)=0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any items in the box), then what is the probability of selecting an apple ? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box ?

Ex. 2


A rare genetic disease is found. Only one in a million people has the causing gene. The genetic diagnosis for this gene is very accurate, and has 100% recall (no false negatives) and 99.99% precision (only 0.01% false positives). Based on Bayes' theorem, decide whether if you would take the diagnosis or not, and explain the reason.

False positive: an error telling healthy person as sick

False negative: an error telling sick person as healthy

Monty Hall Problem

The Monty Hall Problem using Infer.NET



The simulation interface shows three doors. The first door is red and closed, with a large black question mark on it. The second door is blue and open, revealing a blue sports car inside. The third door is yellow and open, revealing a cartoon goat inside.

Door Color	Contents	P (Car Inside)
Red	Unknown (Question Mark)	0.3333
Blue	Car	0.6667
Yellow	Goat	0.0000

Start again

Clear Stats

Success: 20 out of 29
(69% correct)

Appendix: terminology

[illegible]