Support Vector Machine

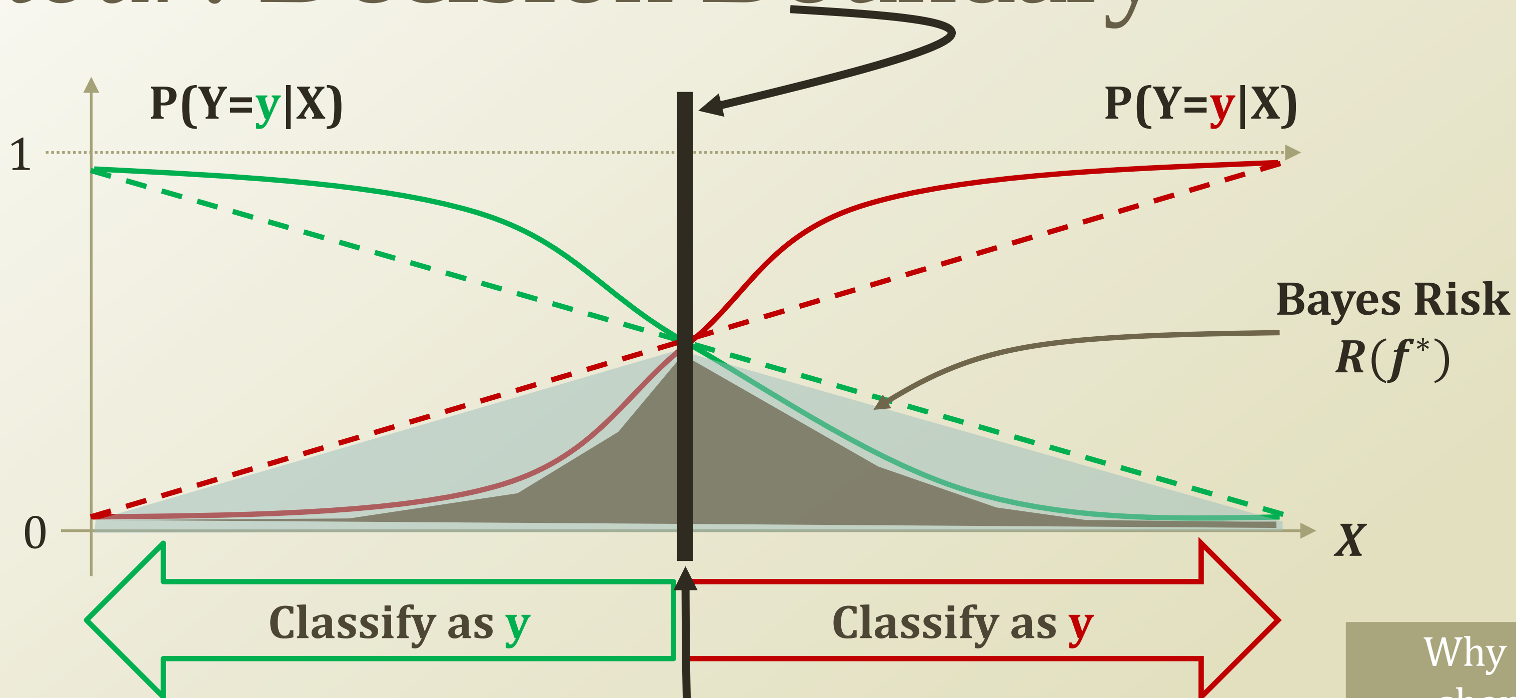
Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

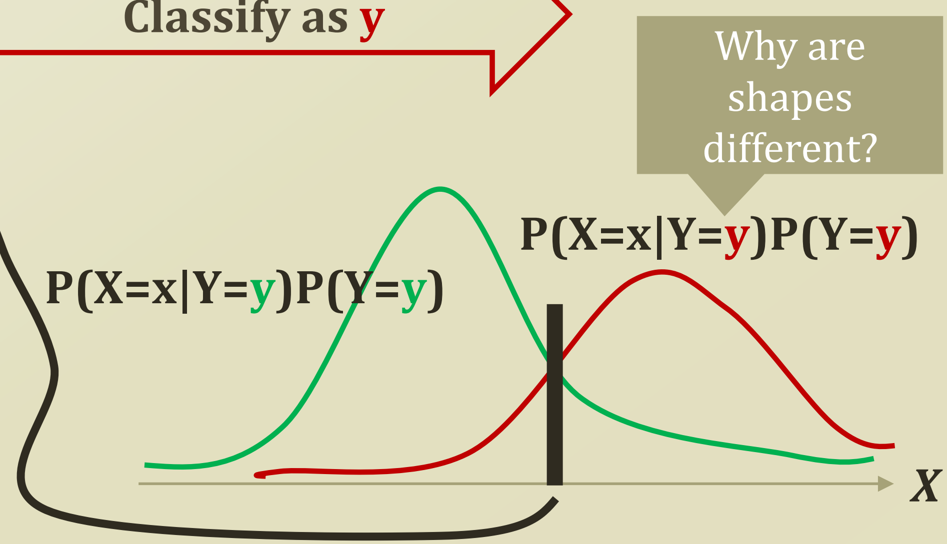
- Learn the support vector machine classifier
  - Understand the maximum margin idea of the SVM
  - Understand the formulation of the optimization problem
- Learn the soft-margin and penalization
  - Know how to add the penalization term
  - Understand the difference between the log-loss and the hinge-loss
- Learn the kernel trick
  - Understand the primal problem and the dual problem of SVM
  - Know the types of kernels
  - Understand how to apply the kernel trick to SVM and logistic regression

SUPPORT VECTOR MACHINE

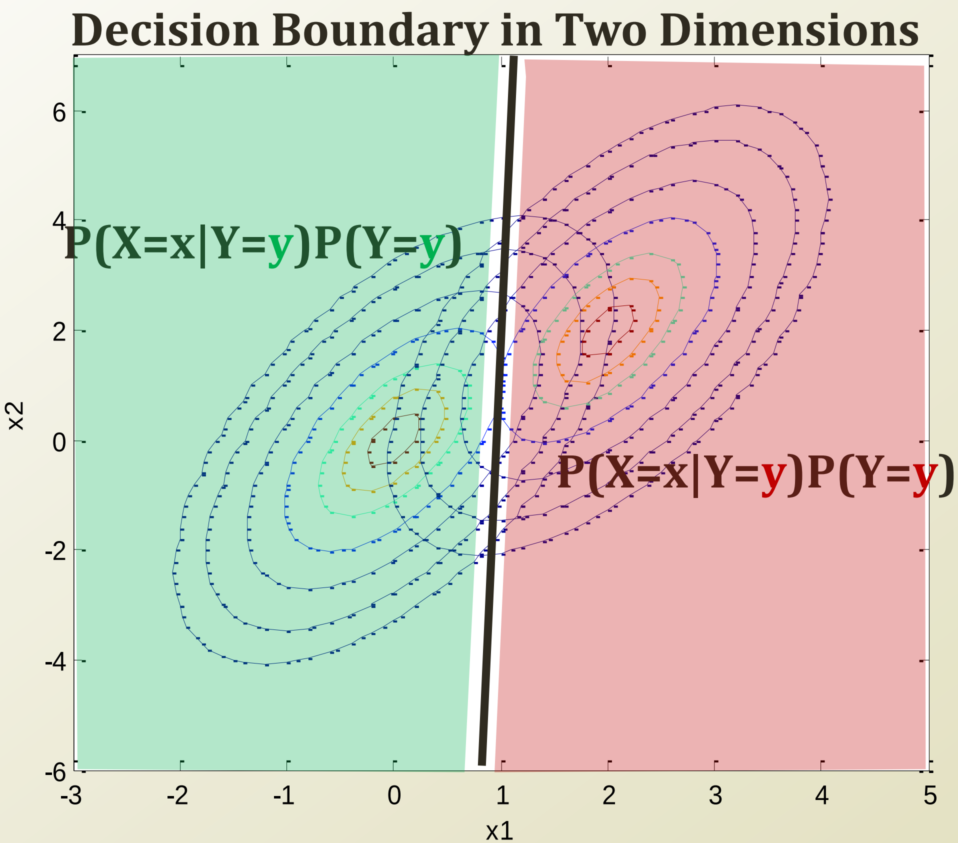
Detour: Decision Boundary



- $f^*(x) = argmax_{Y=y} P(Y = y|X = x)$
  $= argmax_{Y=y} P(X = x|Y = y)P(Y = y)$
- What-if Gaussian class conditional density?
- $P(X = x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



Detour: Decision Boundary in Two Dimension



$$f^*(x) = argmax_{Y=y} P(Y = y|X = x)$$
$$= argmax_{Y=y} P(X = x|Y = y)P(Y = y)$$

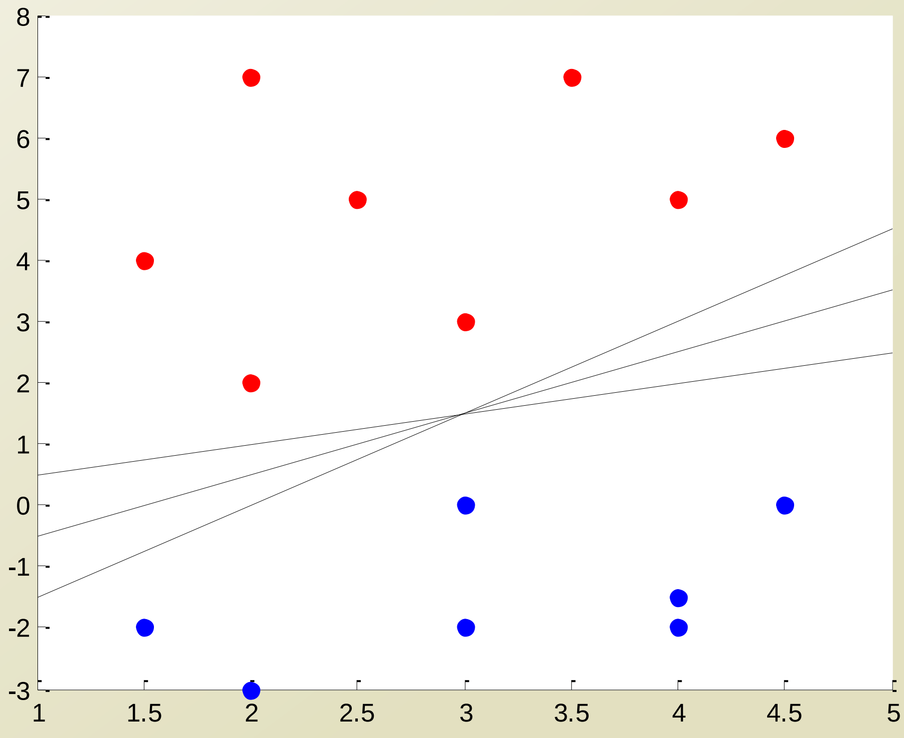
- Two multivariate normal distribution for the class conditional densities
- Decision boundary
  - A linear line
- Linear decision boundary
- Any problem in the real world applications?
  - Observing the combination of x_1 and x_2

$$P(X = x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$P(X = (x_1, x_2)|Y = y) = \frac{1}{\sqrt{2\pi|\Sigma_y|}} \exp(-\frac{(x-\mu_y)\Sigma_y^{-1}(x-\mu_y)'}{2})$$

Decision Boundary without Prob.

- Which is a better decision boundary?
  - Without considering the probability distribution?
- Which points are at the front line?

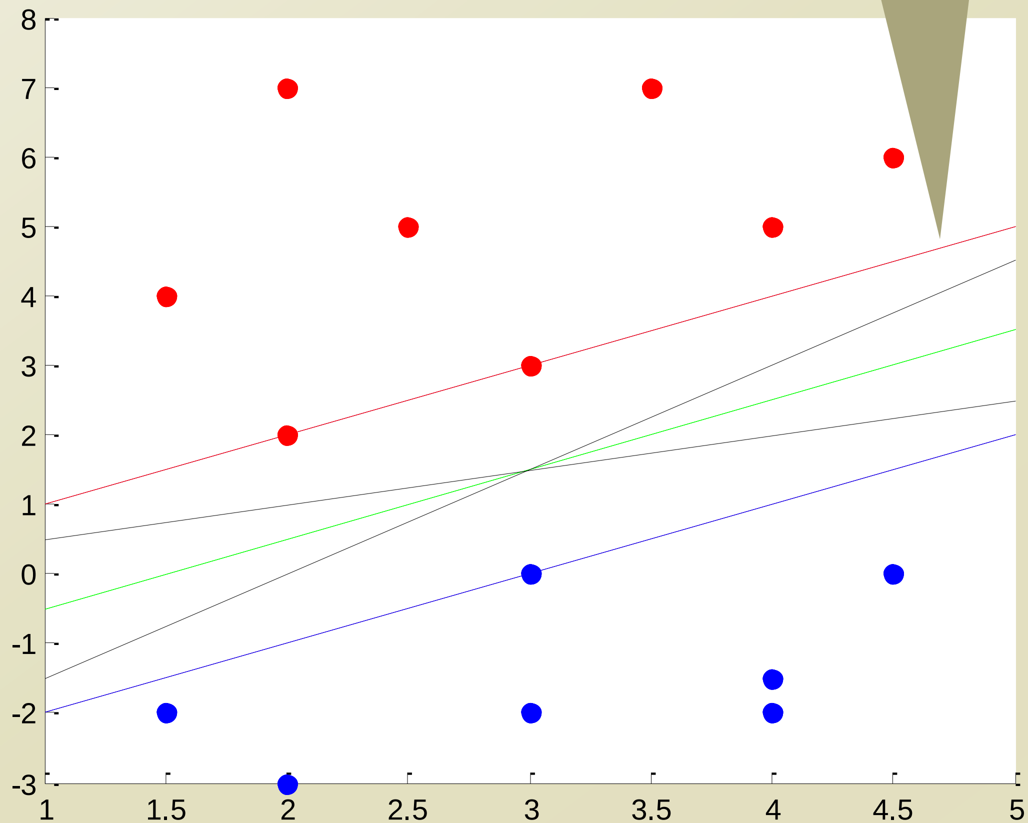


KAIST

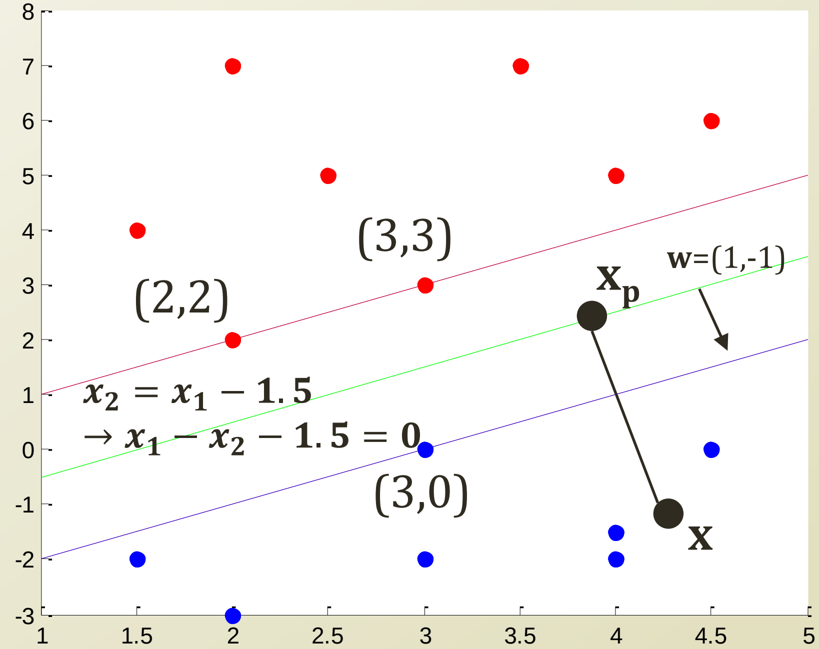
Decision Boundary with Margin

- Decision boundary with maximum margin
  - Between the points close to the boundary
  - How many points?
- Decision boundary line
  - $\mathbf{w} \cdot \mathbf{x} + b = 0$
  - Positive case
    - $\mathbf{w} \cdot \mathbf{x} + b > 0$
  - Negative case
    - $\mathbf{w} \cdot \mathbf{x} + b < 0$
  - Confidence level
    - $(\mathbf{w} \cdot \mathbf{x_j} + b)y_j$
- Margin?
  - Perpendicular distance from the closest point to the decision boundary

How many parameters?



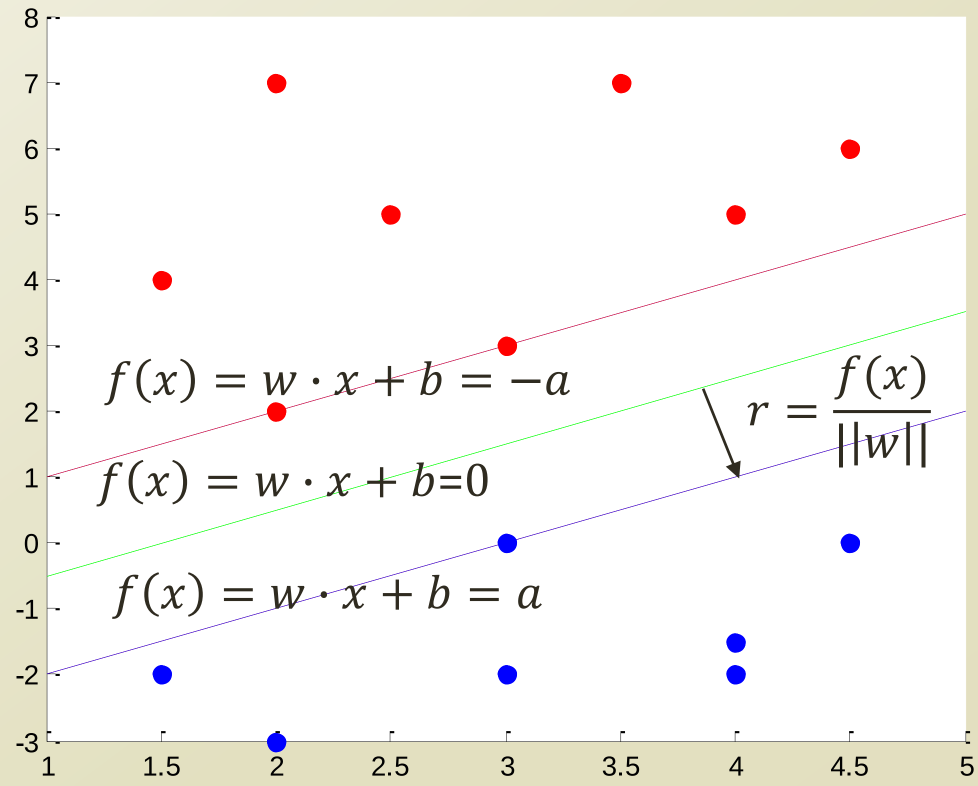
Margin Distance



- Let's say
  - $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$
  - A point $\mathbf{x}$ on the boundary has
    - $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0$
  - A positive point $\mathbf{x}$ has
    - $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = a, a > 0$
- We are going to measure the distance
  - between an arbitrary point $\mathbf{x}$ and a point $\mathbf{x_p}$ on the boundary and on the perpendicular line from $\mathbf{x}$ to the boundary
  - $x = x_p + r\frac{w}{||w||}, f(x_p) = 0$
  - $f(x) = w \cdot x + b = w\left(x_p + r\frac{w}{||w||}\right) + b = \overline{wx_p + b} + r\frac{w \cdot w}{||w||} = r||w||$
- The distance is $r = \frac{f(x)}{||w||}$

Maximizing the Margin

- Good decision boundary?
  - Maximum margin!
  - $r = \frac{a}{||w||}$
  - Need to consider the both side
- Optimization problem?
  - $max_{w,b} 2r = \frac{2a}{||w||}$
    $s.t. (wx_j + b)y_j \geq a, \forall j$
- a is an arbitrary number and can be normalized
  - $min_{w,b} ||w||$
    $s.t. (wx_j + b)y_j \geq 1, \forall j$



This becomes a quadratic optimization problem. Why?

Support Vector Machine with Hard Margin

- Support Vector Machine (SVM)
  - Constructs a set of hyperplanes to have the largest distance to the nearest training data point of any class.
- Hard margin
  - No error cases are allowed
  - What If there is an error case?
- Let's implement the hard margin SVM
  - $min_{w,b}||w||$
    $s.t.\left(wx_j+b\right)y_j \geq 1, \forall j$