Support Vector Machine

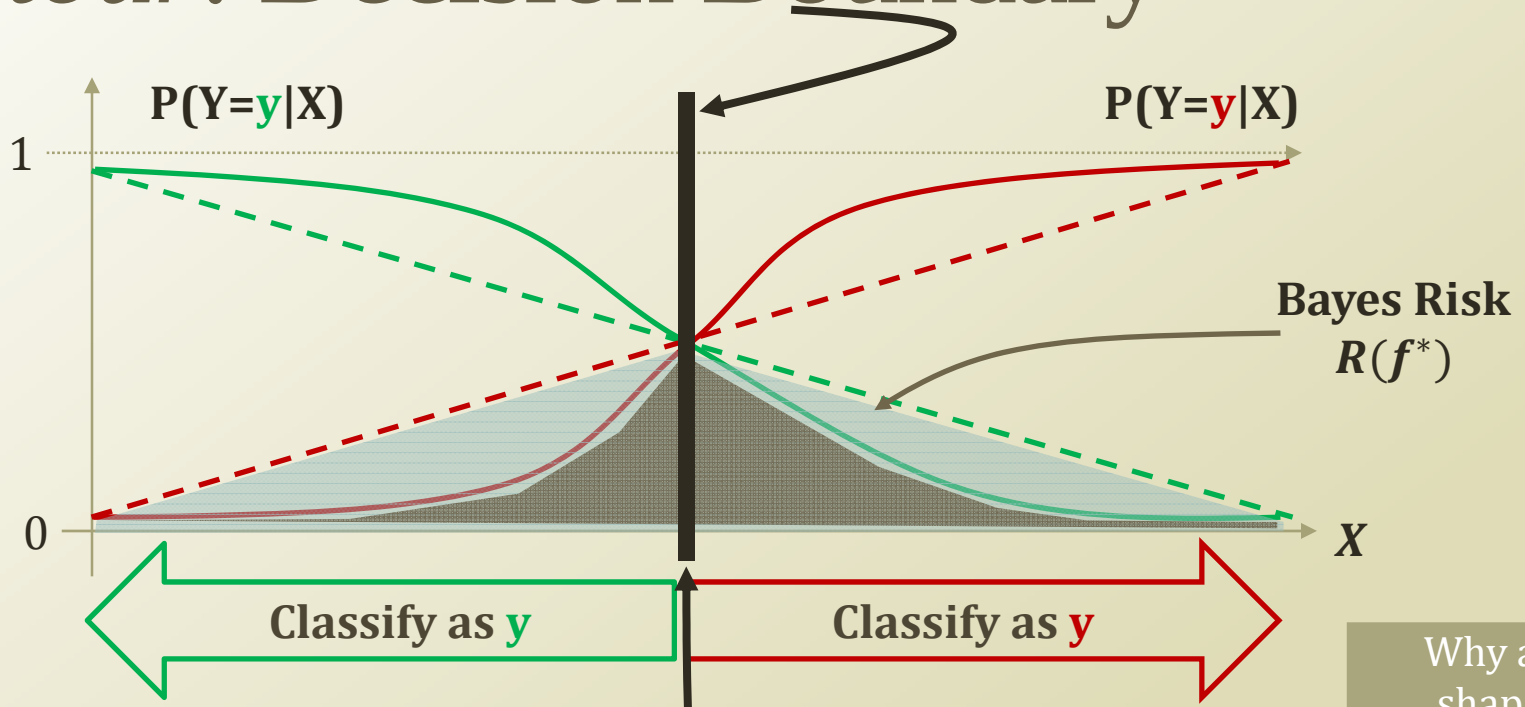
Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

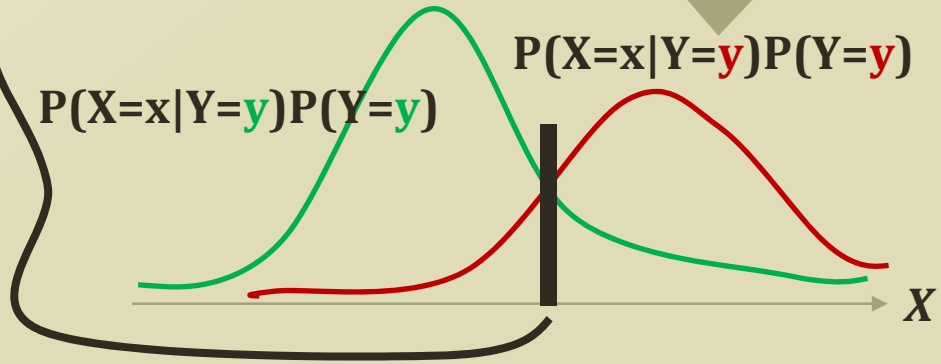
- Learn the support vector machine classifier
  - Understand the maximum margin idea of the SVM
  - Understand the formulation of the optimization problem
- Learn the soft-margin and penalization
  - Know how to add the penalization term
  - Understand the difference between the log-loss and the hinge-loss
- 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

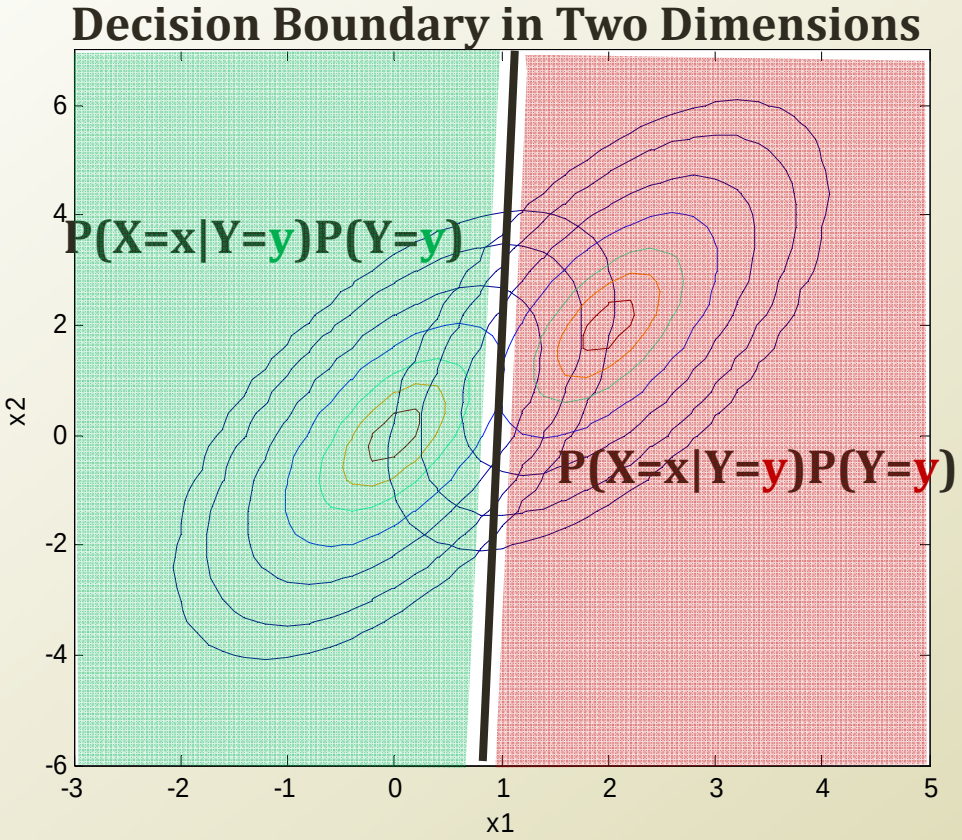


Why are shapes different?



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



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

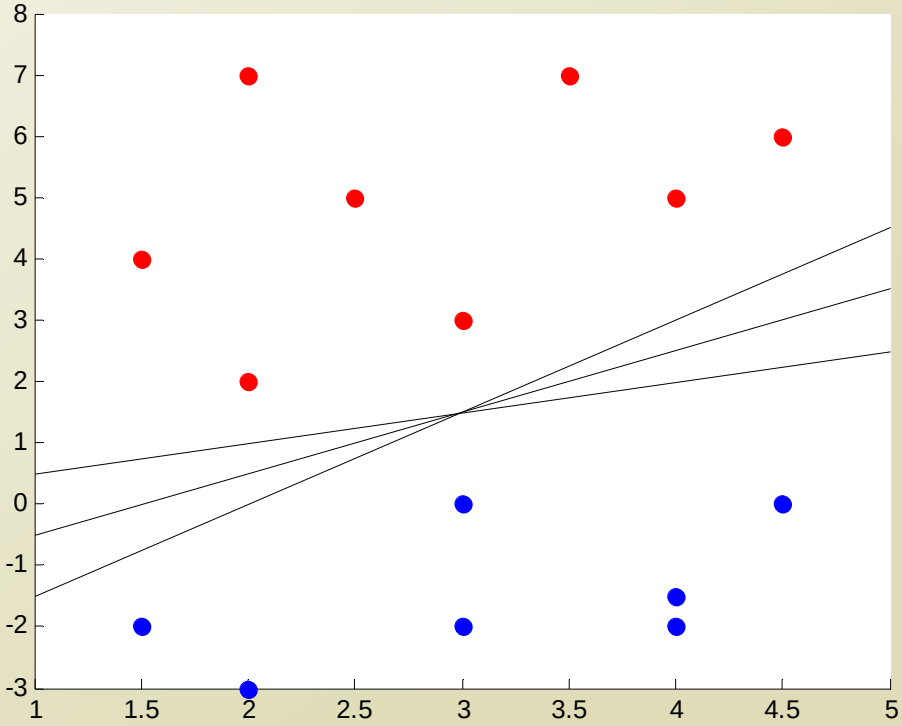
- Two multivariate normal distribution for the class conditional densities
- Decision boundary
  - A linear line
- Linear decision boundary
- Any problem in the real world applications?
  - Observing the combination of x_1 and x_2

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

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

Decision Boundary without Prob.

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



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