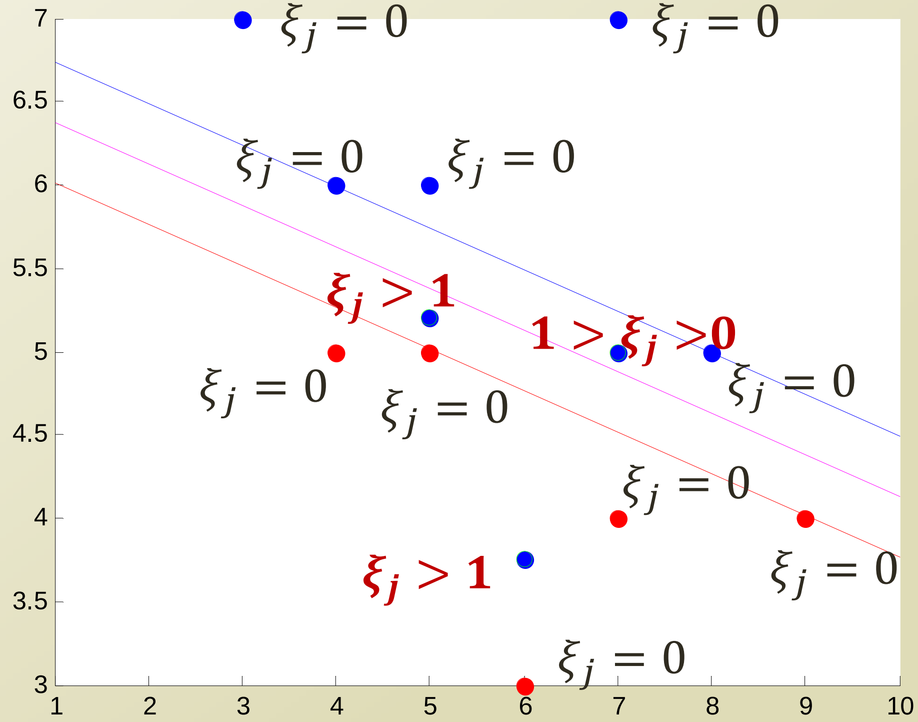
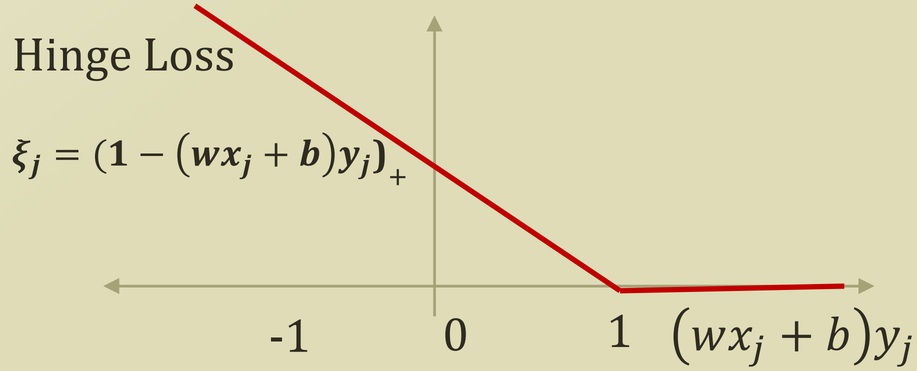
Support Vector Machine

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

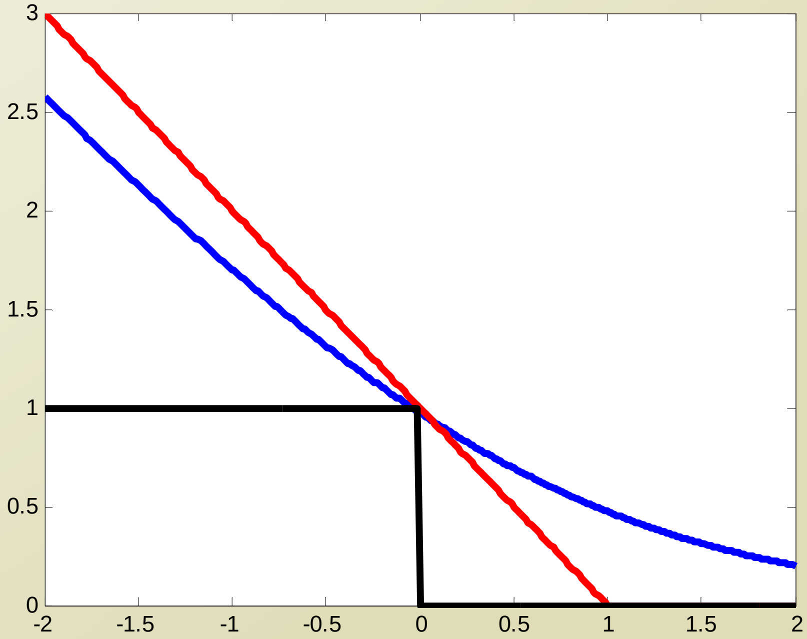
Soft-Margin SVM

- $min_{w,b}||w|| + C\sum_j \xi_j$
  $s.t.$
  $(wx_j + b)y_j \geq 1 - \xi_j, \forall j$
  $\xi_j \geq 0, \forall j$
- We soften the constraints
  - By adding a slack variable
- Instead, we penalize the mis-classification cases in the objective function
  - $C\sum_j \xi_j$
- How to recover the hard-margin SVM?





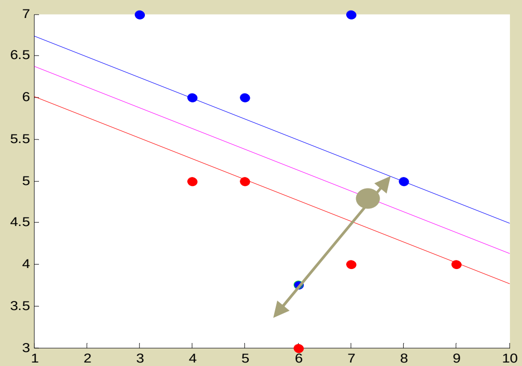
Comparison to Logistic Regression



- Loss function
  - $\xi_j = loss(f(x_j), y_j)$
- SVM loss function: Hinge Loss
  - $\xi_j = (1 - (wx_j + b)y_j)_+$
- Logistic Regression loss function: Log Loss
  - $\hat{\theta} = argmax_\theta \sum_{1 \le i \le N} log(P(Y_i|X_i;\theta))$

$$= argmax_\theta \sum_{1 \le i \le N} \{Y_i X_i \theta - \log(1 + e^{X_i \theta})\}$$

  - $\xi_j = -\log\left(P(Y_j|X_j, w, b)\right) = \log\left(1 + e^{(wx_j + b)y_j}\right)$
- Which loss function is preferable?
  - Around the decision boundary?
  - Overall place?



Strength of the Loss Function

- $min_{w,b,\xi_j} ||w|| + C\sum_j \xi_j$
  $s.t.$
  $(wx_j + b)y_j \geq 1 - \xi_j, \forall j$
  $\xi_j \geq 0, \forall j$
- Let's implement the model
- How does the decision boundary evolves over the variations of C?