

# Support vector machine

Kookjin Lee

([kookjin.Lee@asu.edu](mailto:kookjin.Lee@asu.edu))

The contents of this course, including lectures and other instructional materials, are copyrighted materials. Students may not share outside the class, including uploading, selling or distributing course content or notes taken during the conduct of the course. Any recording of class sessions is authorized only for the use of students enrolled in this course during their enrollment in this course. Recordings and excerpts of recordings may not be distributed to others. (see ACD 304 – 06 , “ Commercial Note Taking Services ” and ABOR Policy 5 - 308 F.14 for more information).

# Going back to logistic regression

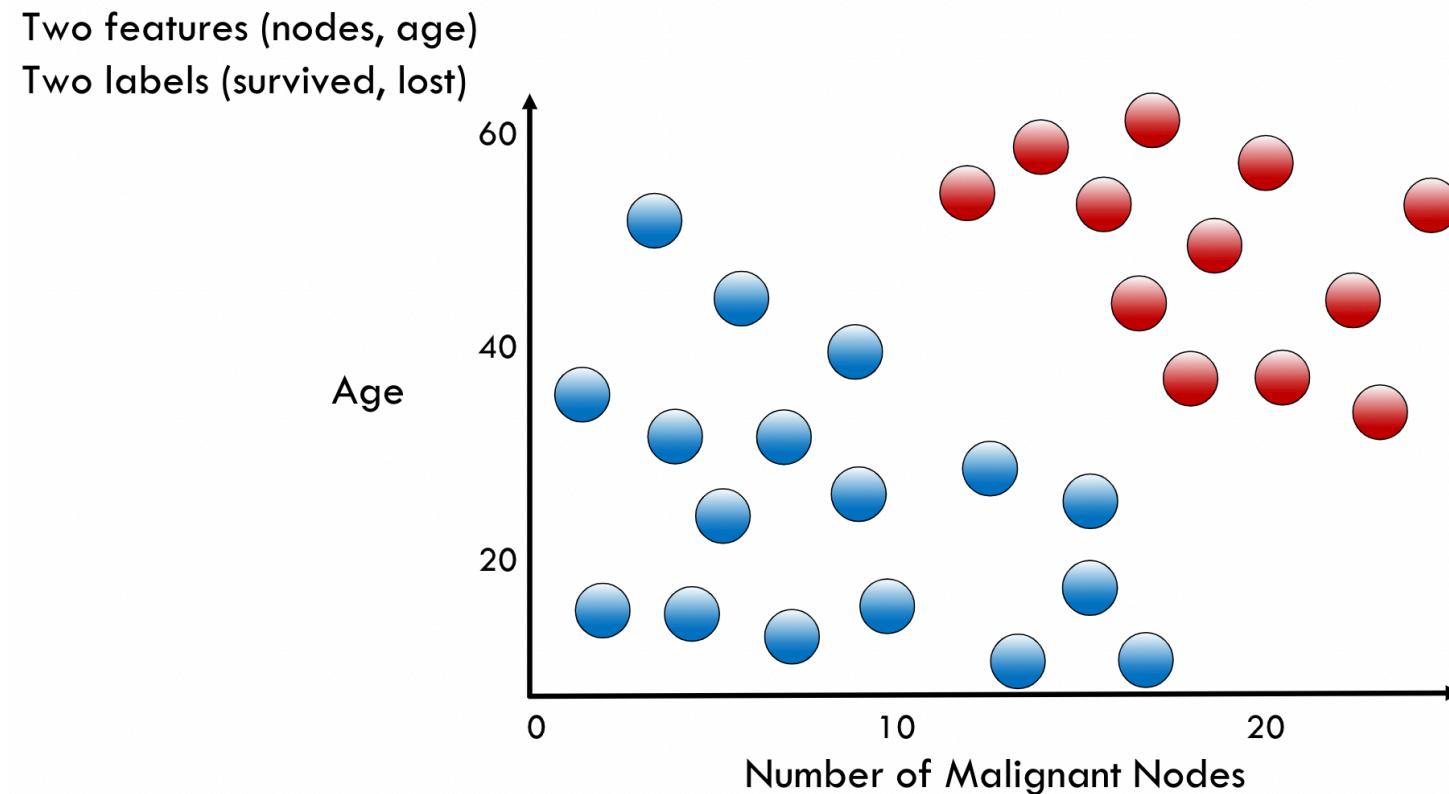
- Logistic regression

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

- Threshold: .5
  - Predict 1 if  $\theta^T x \geq 0$
  - Predict 0 otherwise
- What we hope for:
  - If “1”,  $\theta^T x \gg 0$  and if “0”,  $\theta^T x \ll 0$

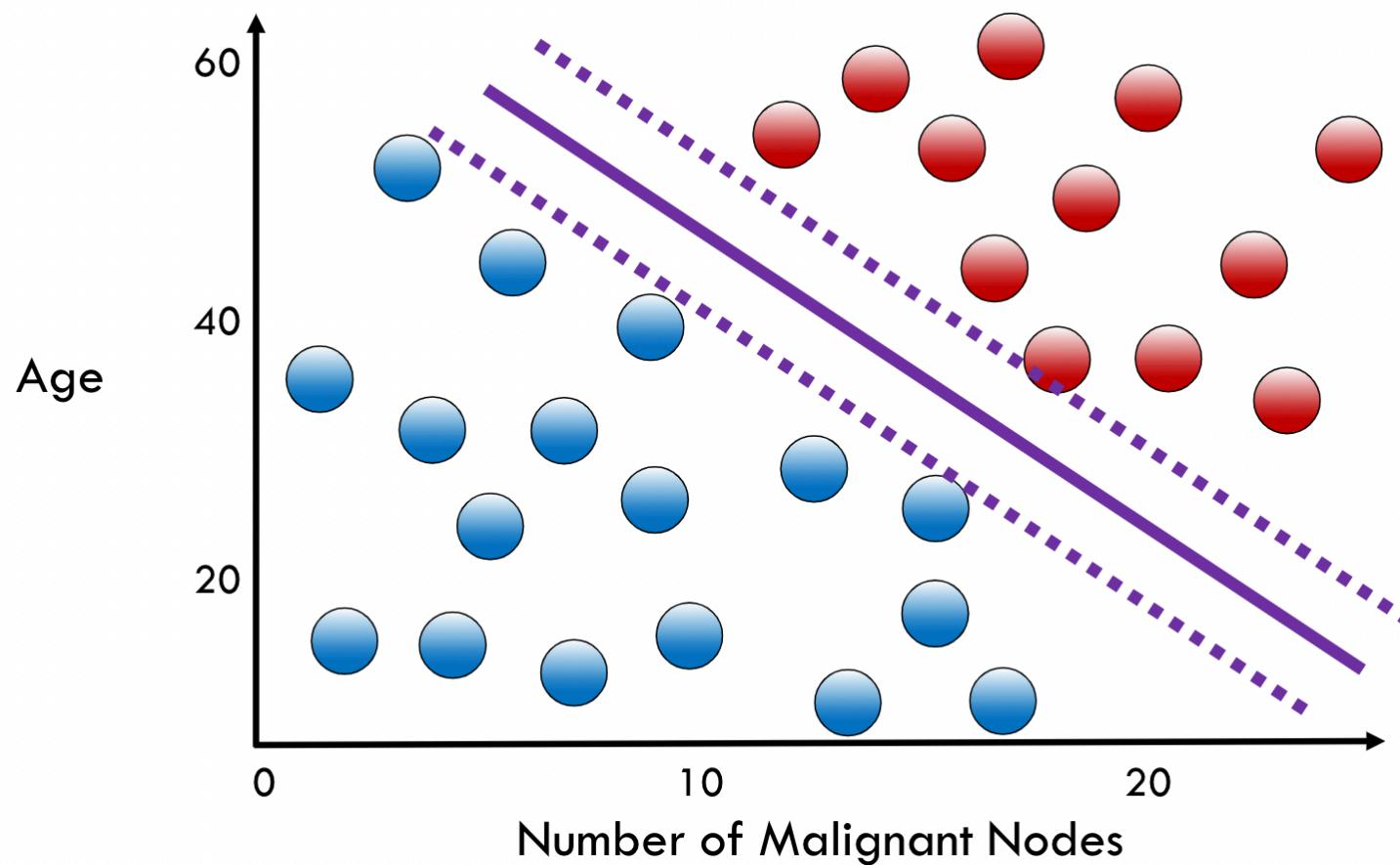
# Margins of separation

- Given a data, we could have different lines:



# Maximal separation margin

- With SVM, we would like to find



# SVM

- Previously in logistic regression:

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

- Class labels are 0 or 1
- In SVM,  
$$h_{w,b}(x) = g(w^T x + b)$$
- Class labels are -1 or 1

# SVM

- Maximizing the (functional) margin

- Margin:

$$\delta^{(i)} = y^{(i)}(\mathbf{w}^\top \mathbf{x}^{(i)} + b)$$

- We hope

- If “1”,

$$\mathbf{w}^\top \mathbf{x}^{(i)} + b \gg 0$$

- If “-1”,

$$\mathbf{w}^\top \mathbf{x}^{(i)} + b \ll 0$$

- All together,

$$\delta^{(i)} \gg 0$$

# SVM – continued

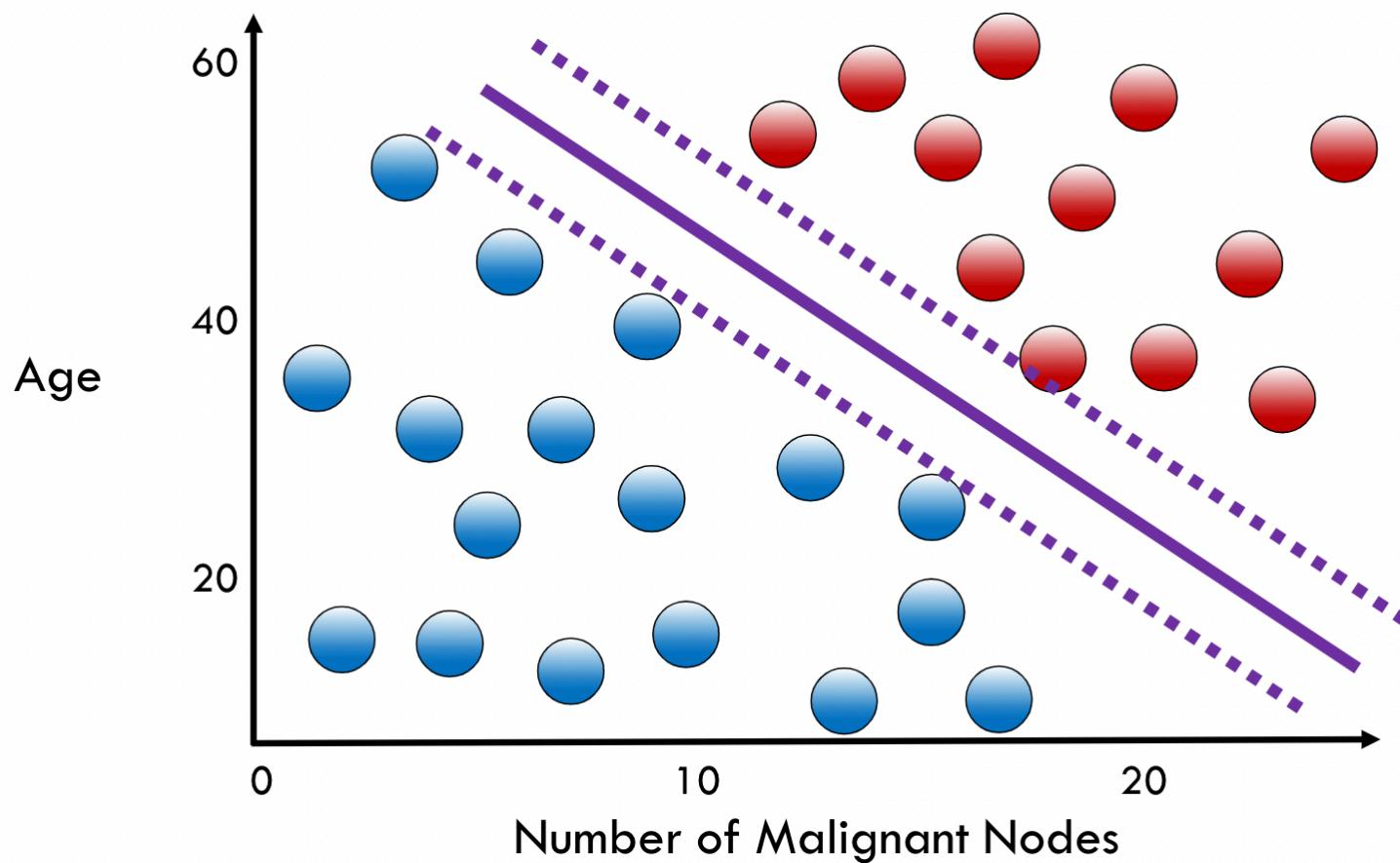
- Functional margin w.r.t. training set

$$\min \delta^{(i)}$$

- Normalization:

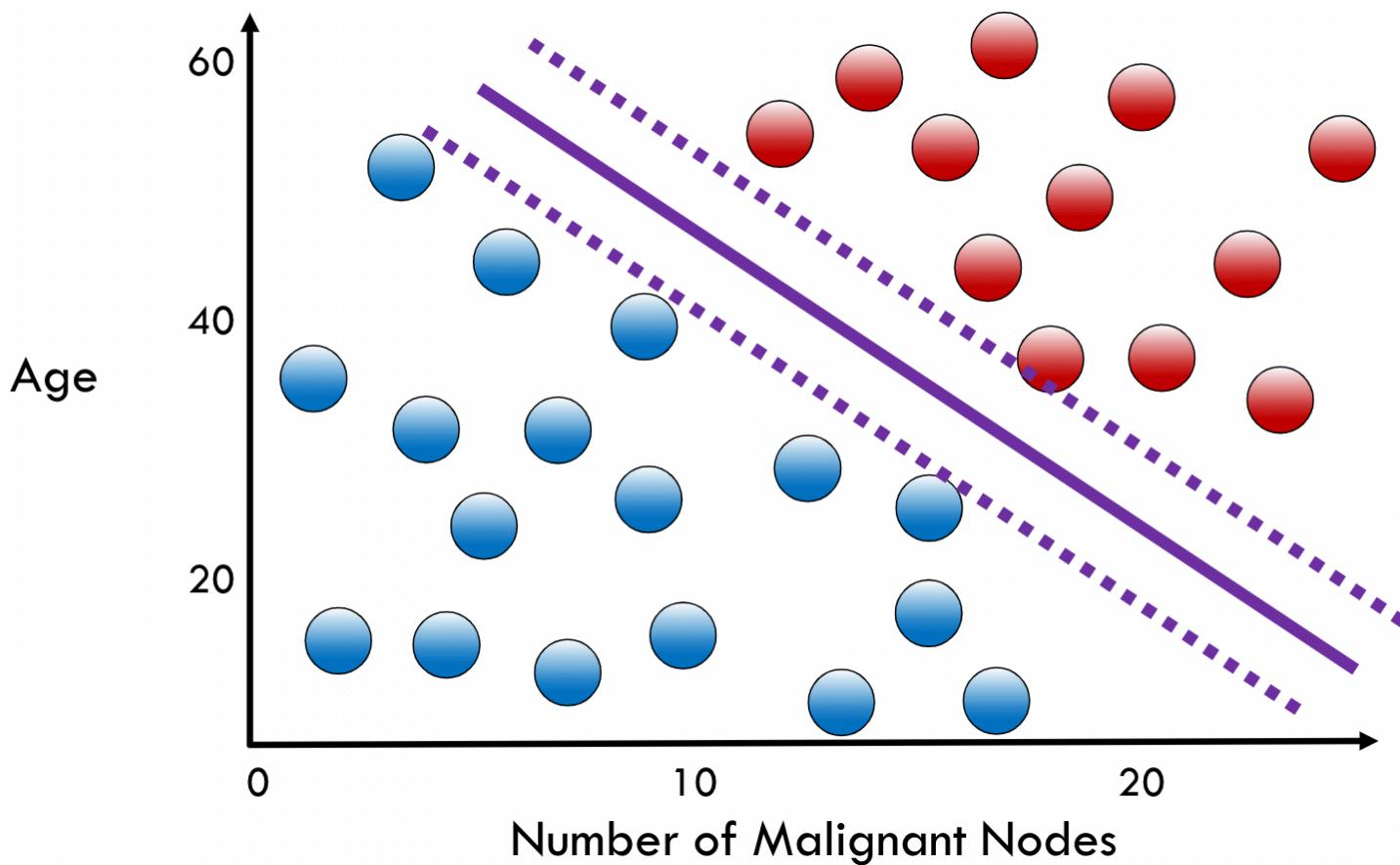
$$\frac{w}{\|w\|_2} \quad \frac{b}{\|w\|_2}$$

# SVM – geometric margin



# SVM – geometric margin

- Euclidean distance from a data instance



$$\delta^{(i)} = \frac{\mathbf{w}^\top \mathbf{x}^{(i)} + b}{\|\mathbf{w}\|_2}$$

$$\mathbf{w}^\top \left( \mathbf{x}^{(i)} - \frac{\mathbf{w}}{\|\mathbf{w}\|_2} \delta^{(i)} \right) + b = 0$$

$$\delta^{(i)} = \frac{y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b)}{\|\mathbf{w}\|_2}$$

# SVM – maximizing the margin

- Optimization problem:

$$\max_{\mathbf{w}, b} \delta$$

such that  $\frac{y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b)}{\|\mathbf{w}\|_2} \geq \delta$

# SVM – (informal) derivation

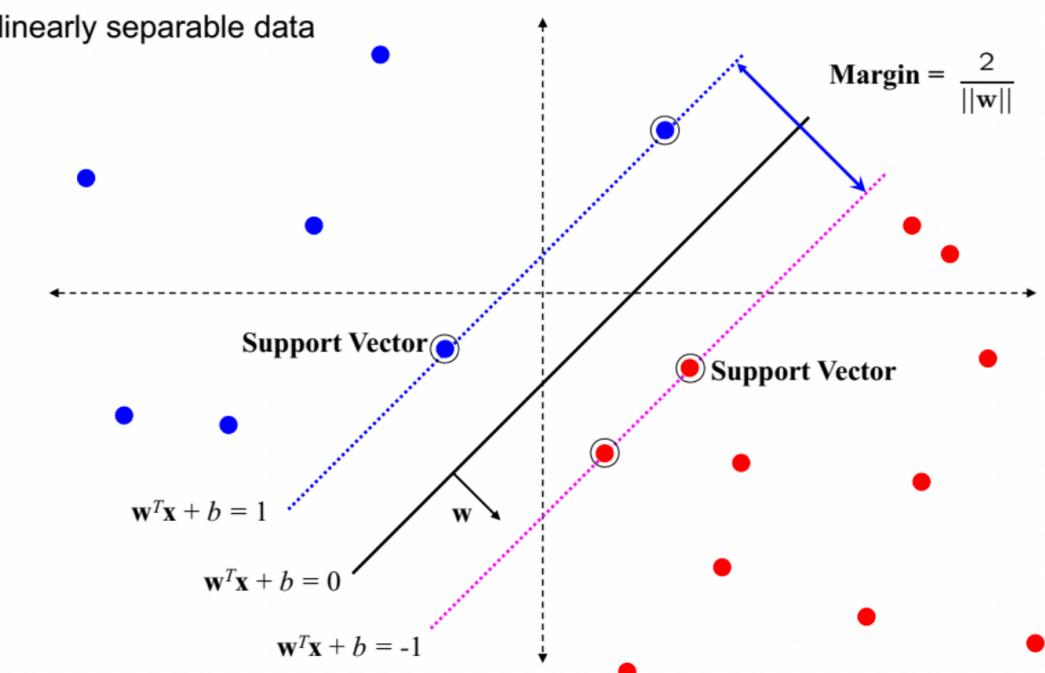
- Choose normalization such that

- for positive class,  $\mathbf{w}^\top \mathbf{x}^{(i)} + b = 1$
- for negative class,  $\mathbf{w}^\top \mathbf{x}^{(i)} + b = -1$

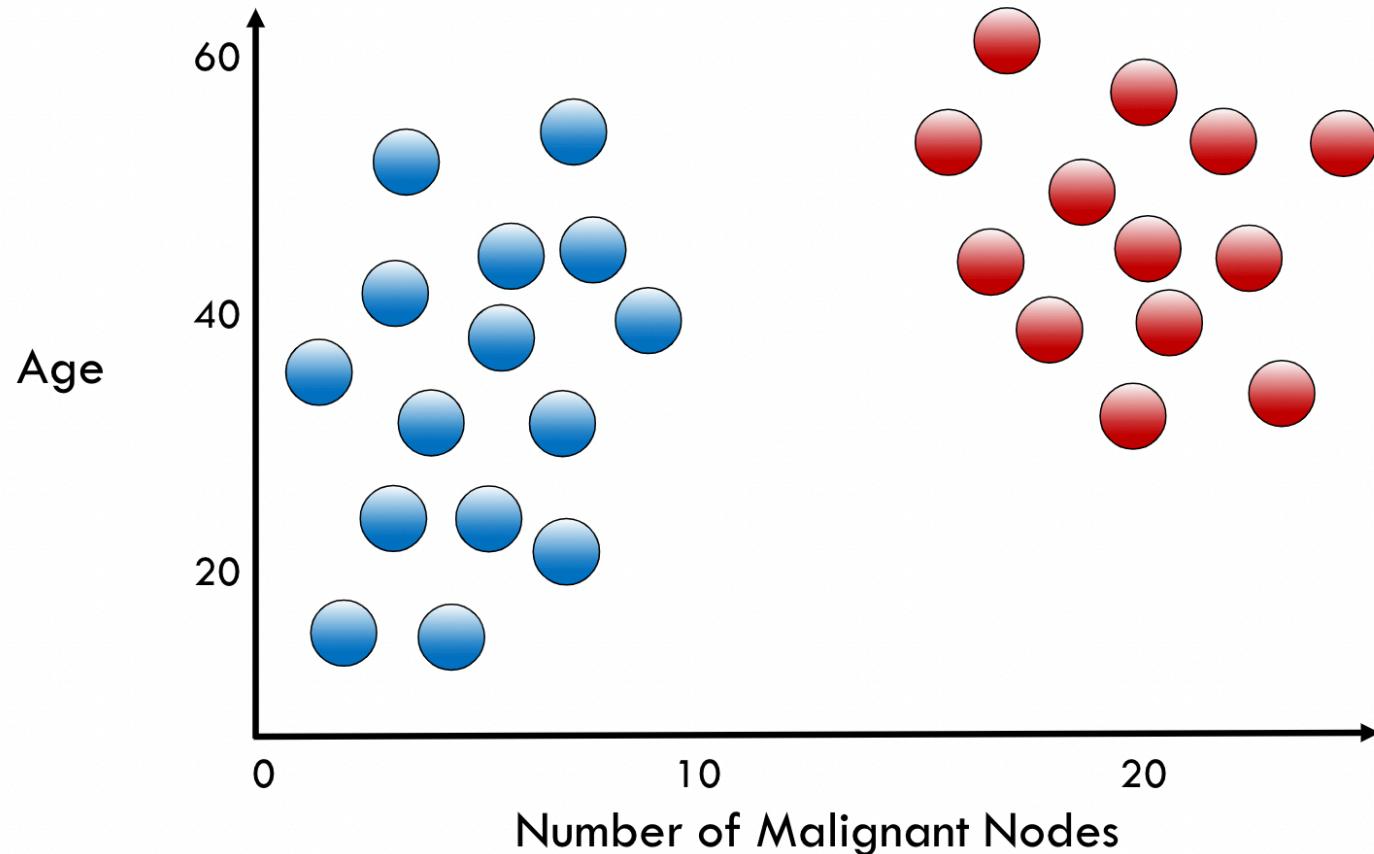
- Optimization objective:

$$\max \frac{2}{\|\mathbf{w}\|_2} \quad \text{s.t.} \quad \mathbf{w}^\top \mathbf{x}^{(i)} + b \geq 1 \text{ if } y^{(i)} = 1$$
$$\mathbf{w}^\top \mathbf{x}^{(i)} + b \leq -1 \text{ if } y^{(i)} = -1$$

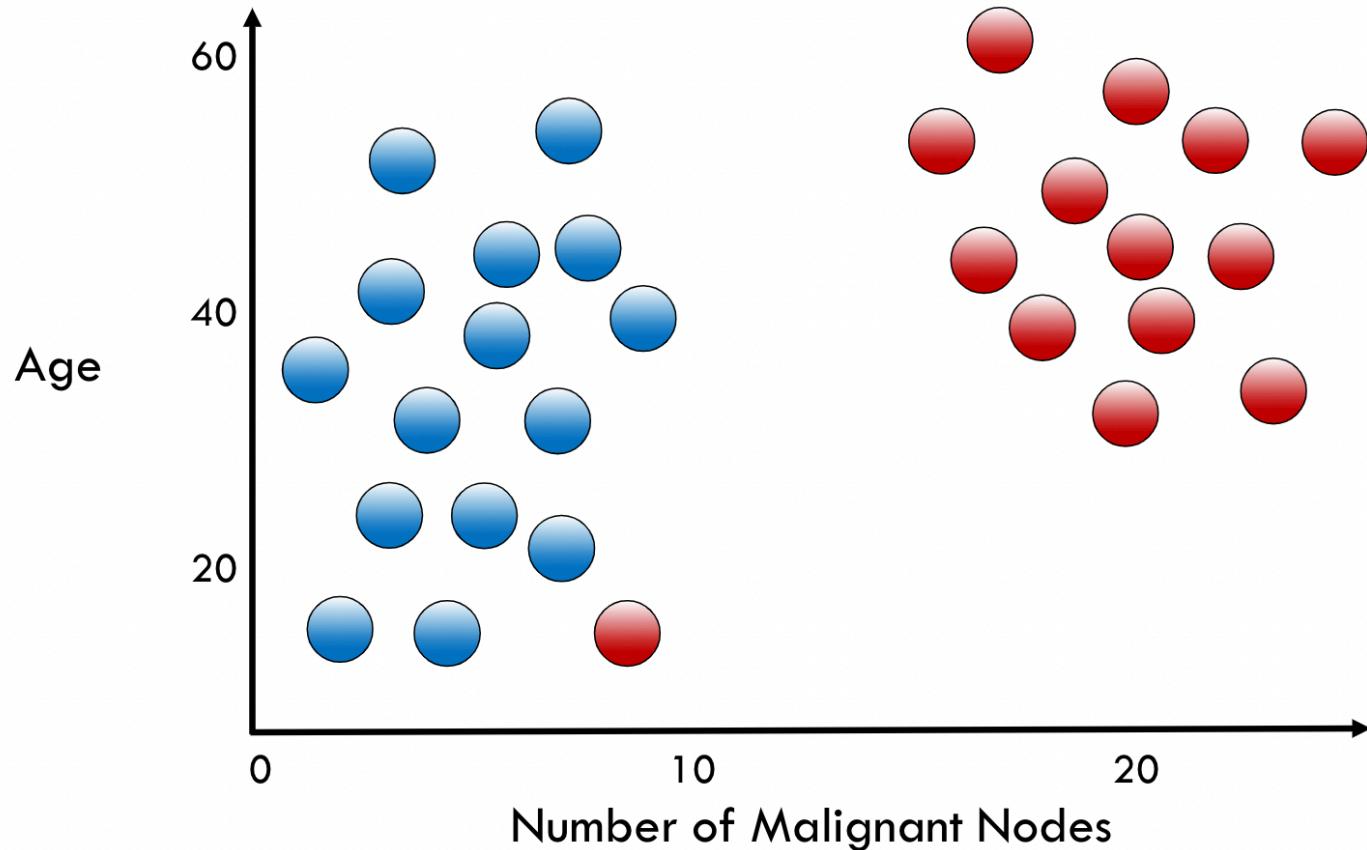
$$\min \|\mathbf{w}\| \quad \text{s.t.} \quad y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1$$



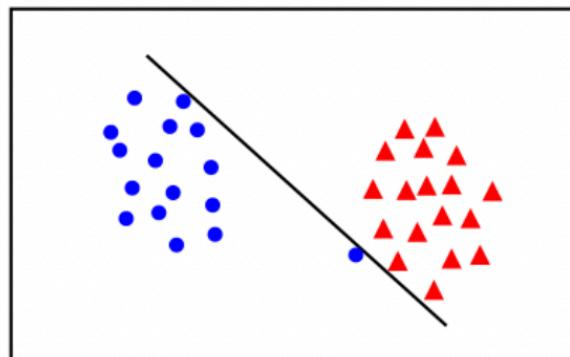
# Outlier sensitivity



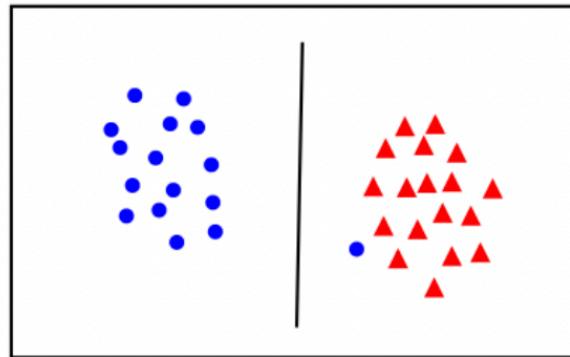
# Outlier sensitivity



# What is the best model parameter?



- the points can be linearly separated but there is a very narrow margin



- but possibly the large margin solution is better, even though one constraint is violated

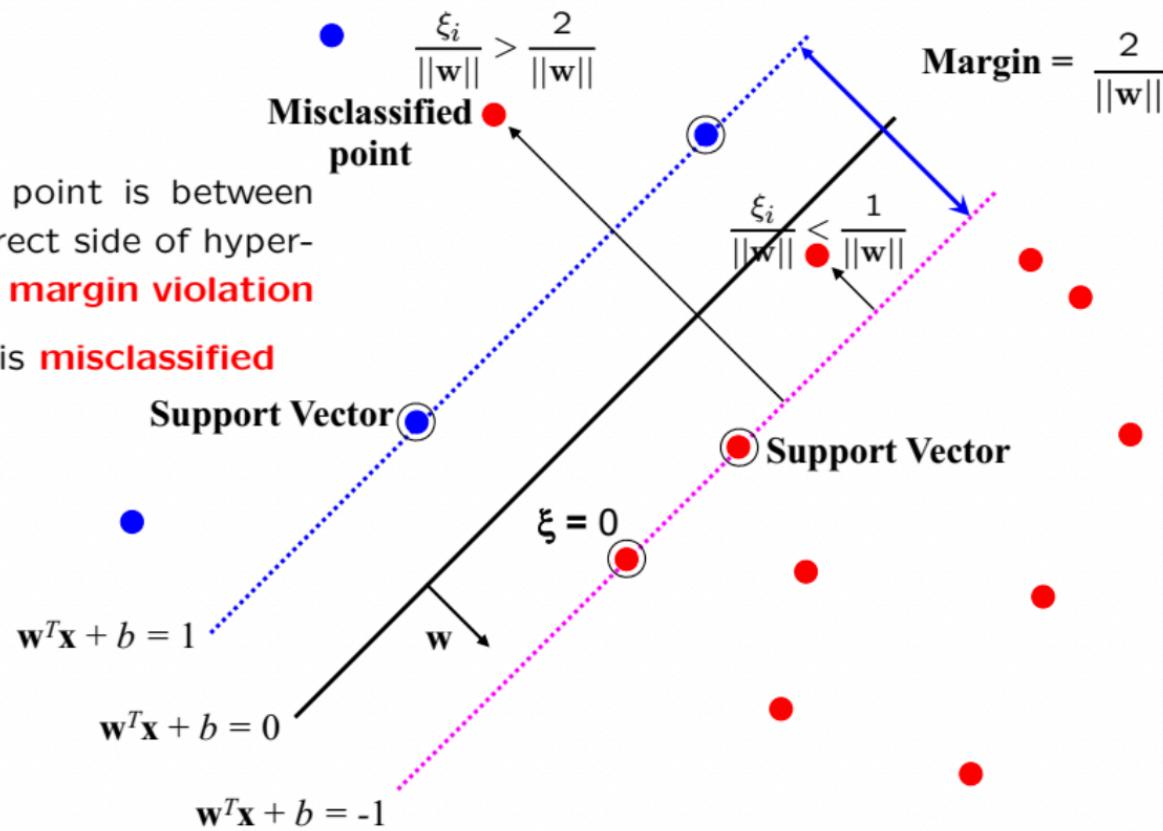
In general there is a trade off between the margin and the number of mistakes on the training data

# Regularization

- Introducing slack variables

$$\xi_i \geq 0$$

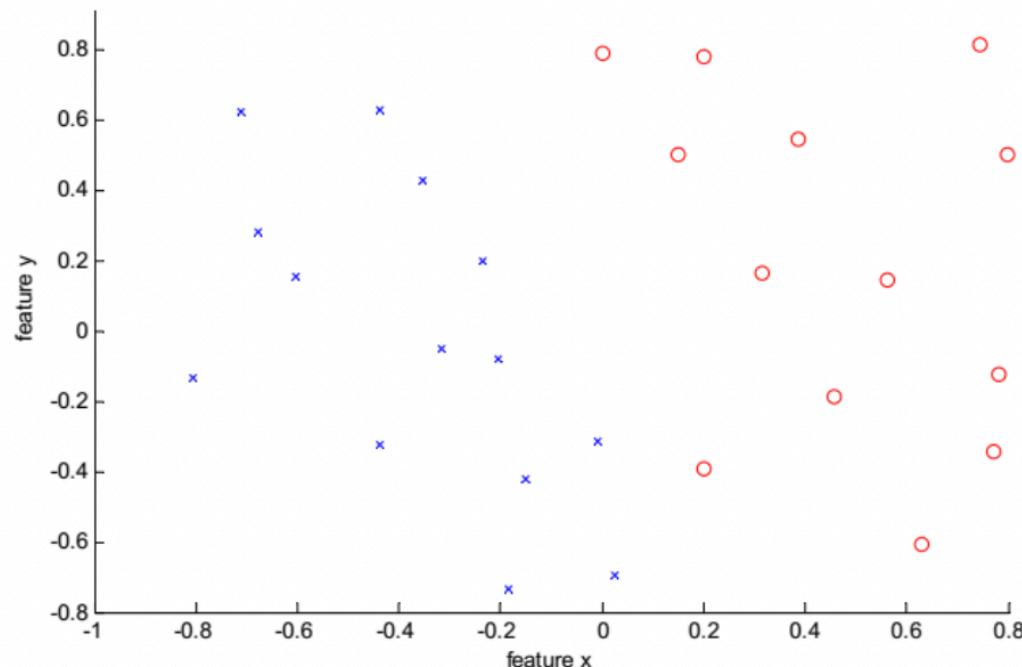
- for  $0 < \xi \leq 1$  point is between margin and correct side of hyperplane. This is a **margin violation**
- for  $\xi > 1$  point is **misclassified**



# “Soft” margin SVM

- A new optimization objective:

$$\min \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad \text{s.t.} \quad y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) \geq 1 - \xi_i$$



# Hard vs soft SVM

- With different values of C

