Support Vector Machines

ECE 4200
2/19/2020

References

Chapter 7, Bishop

http://cs229.stanford.edu/notes/cs229-notes3.pdf

Recall the set-up

Given:

$$S = \{(\bar{X}_1, y_1), (\bar{X}_2, y_2), \ldots, (\bar{X}_n, y_n)\}$$

Interested in linear classifiers: $(\bar{w}, t)$,

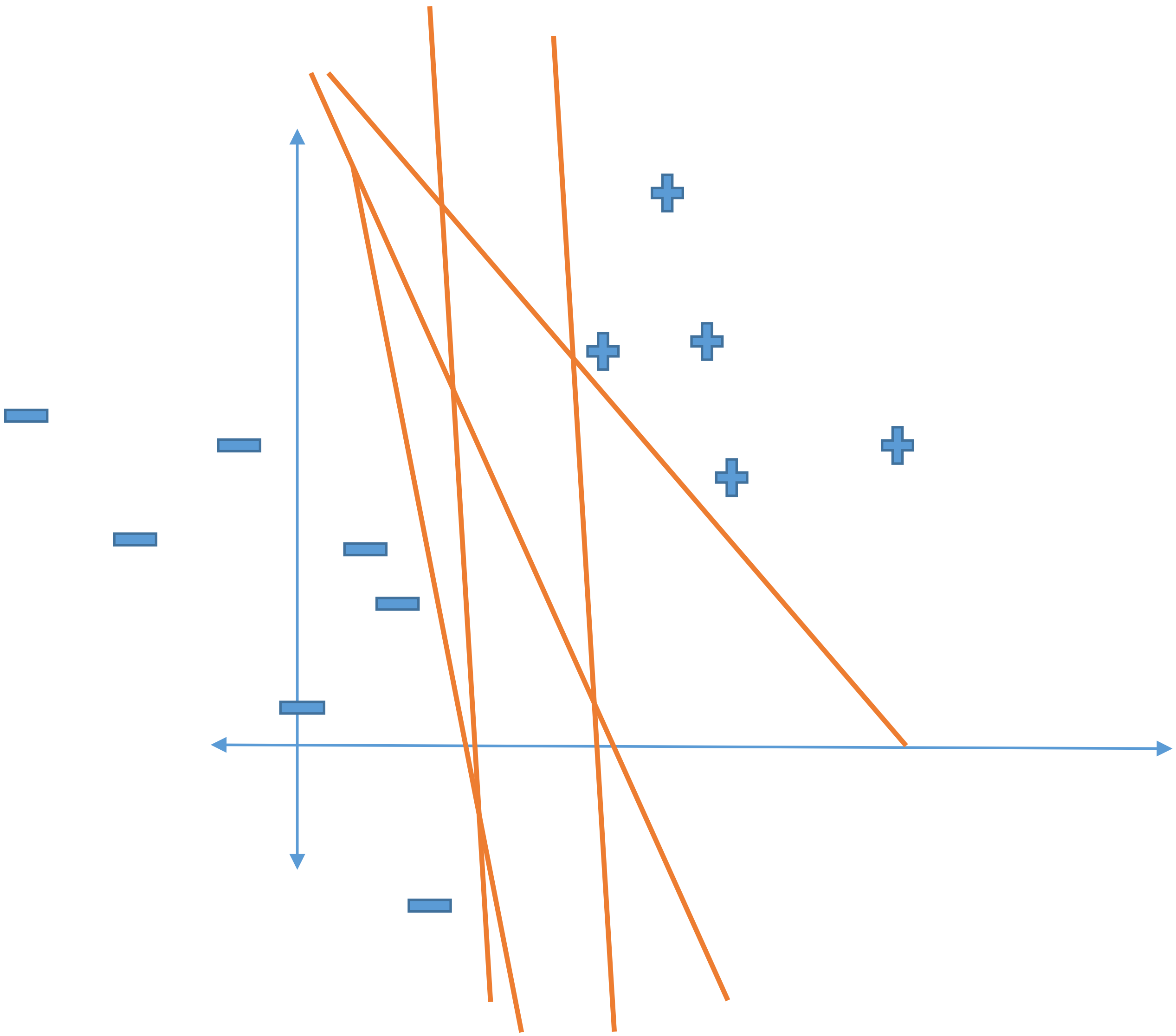
For a new feature $\bar{X} \in R^d$, we output

$$\text{sign}\ (\bar{X} \cdot \bar{w} - t)$$

Maximum margin classifiers

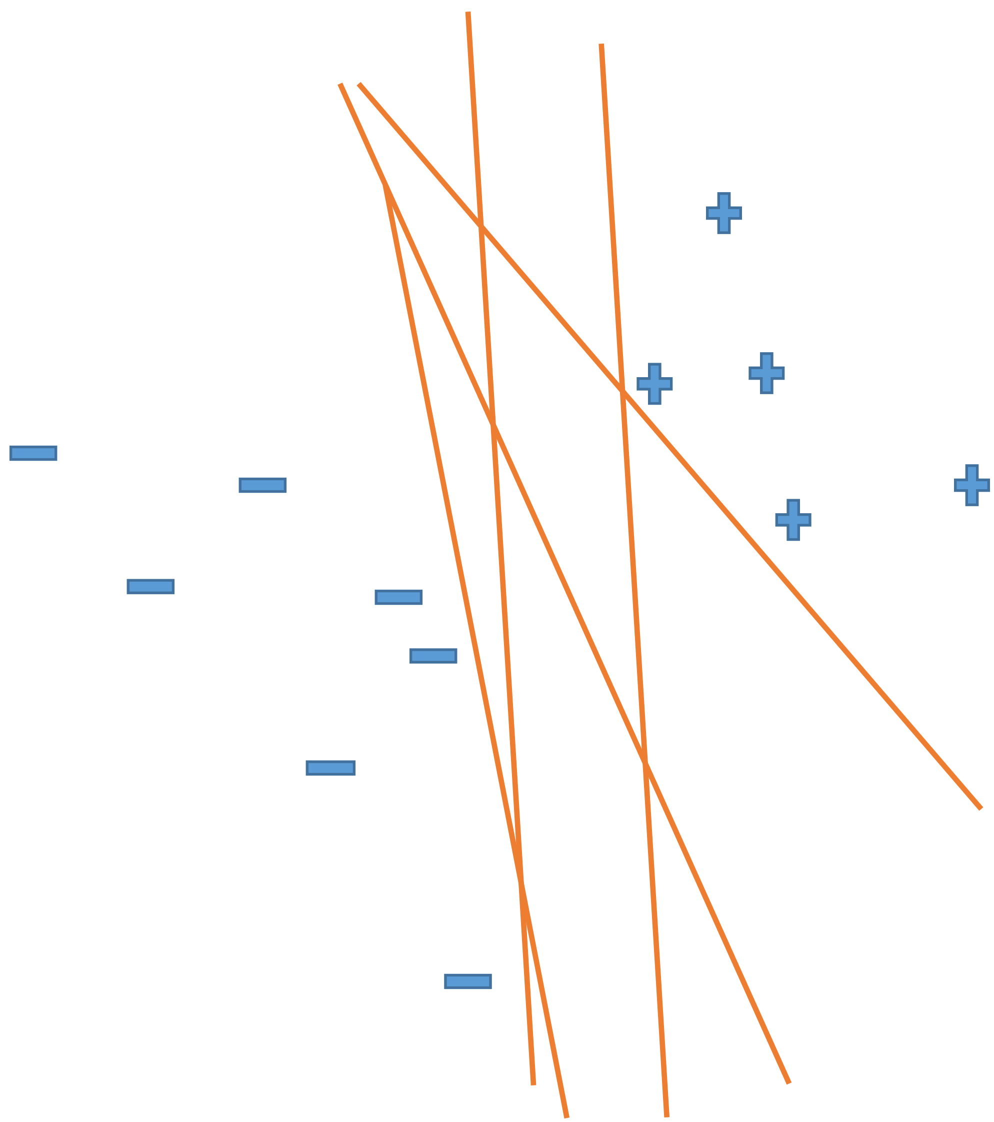
Assume data linearly separable

What is the best linear classifier?



Assume data linearly separable

What is the best linear classifier?



Recall margin

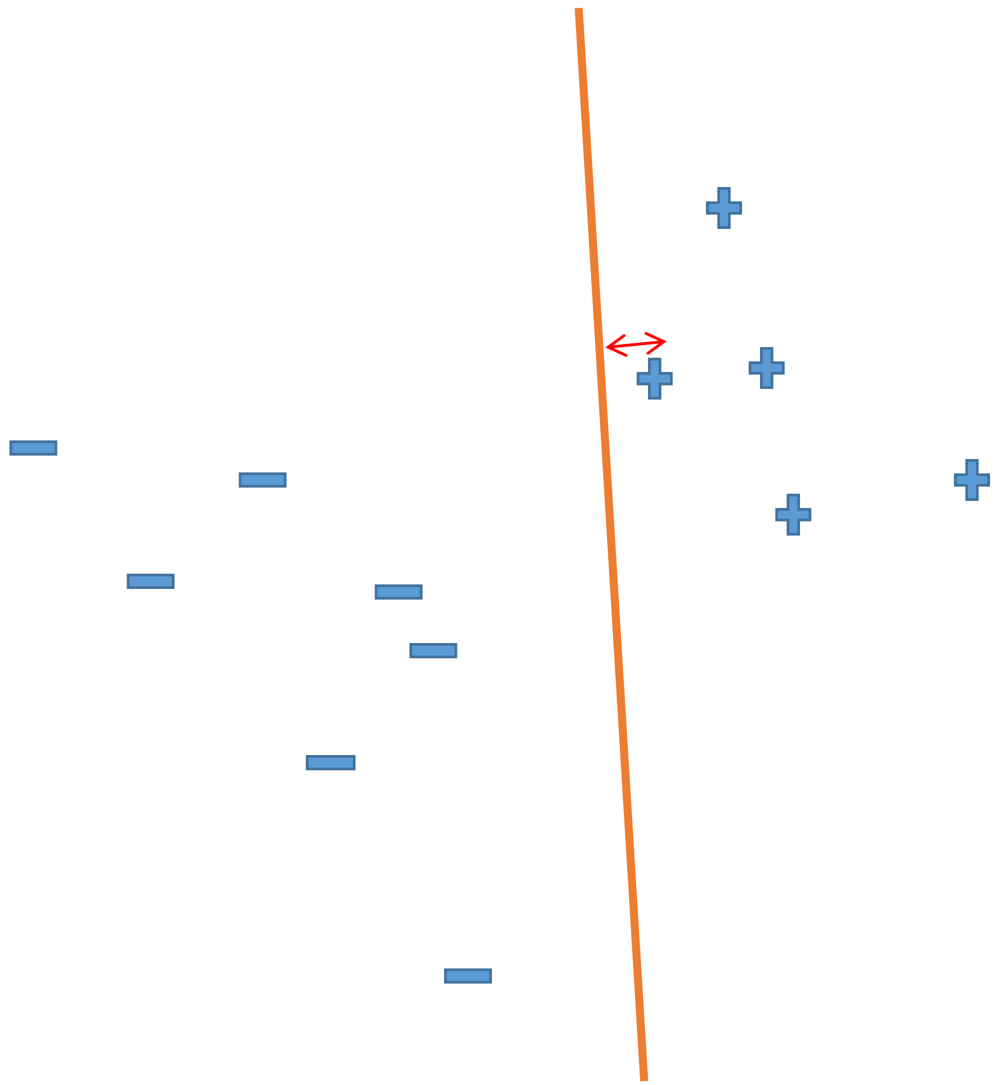
For a hyper plane $(\overline{w}, t)$,

that separates S,

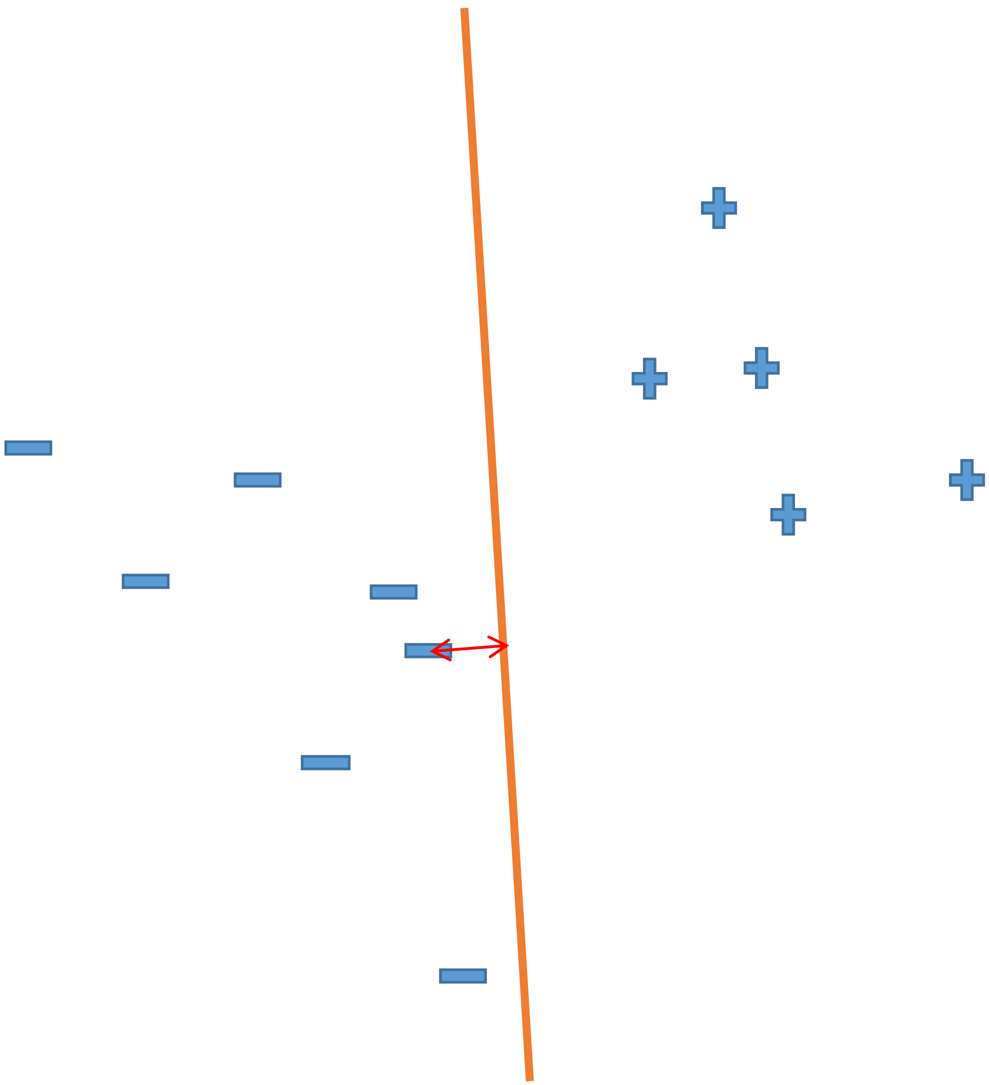
margin is the distance of $(\overline{w}, t)$ to the **closest** point in S.

Margin of S: largest of all margins

Recall margin

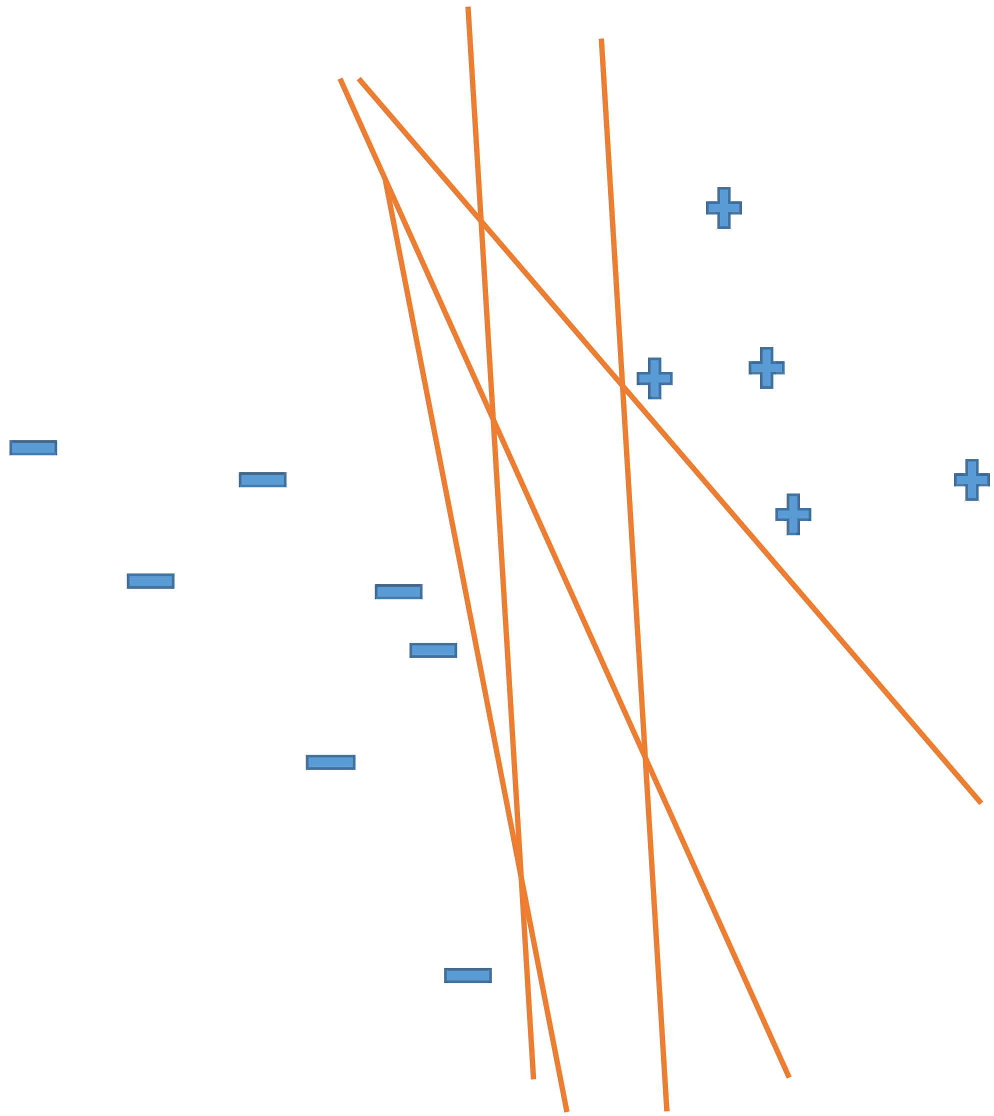


Recall margin



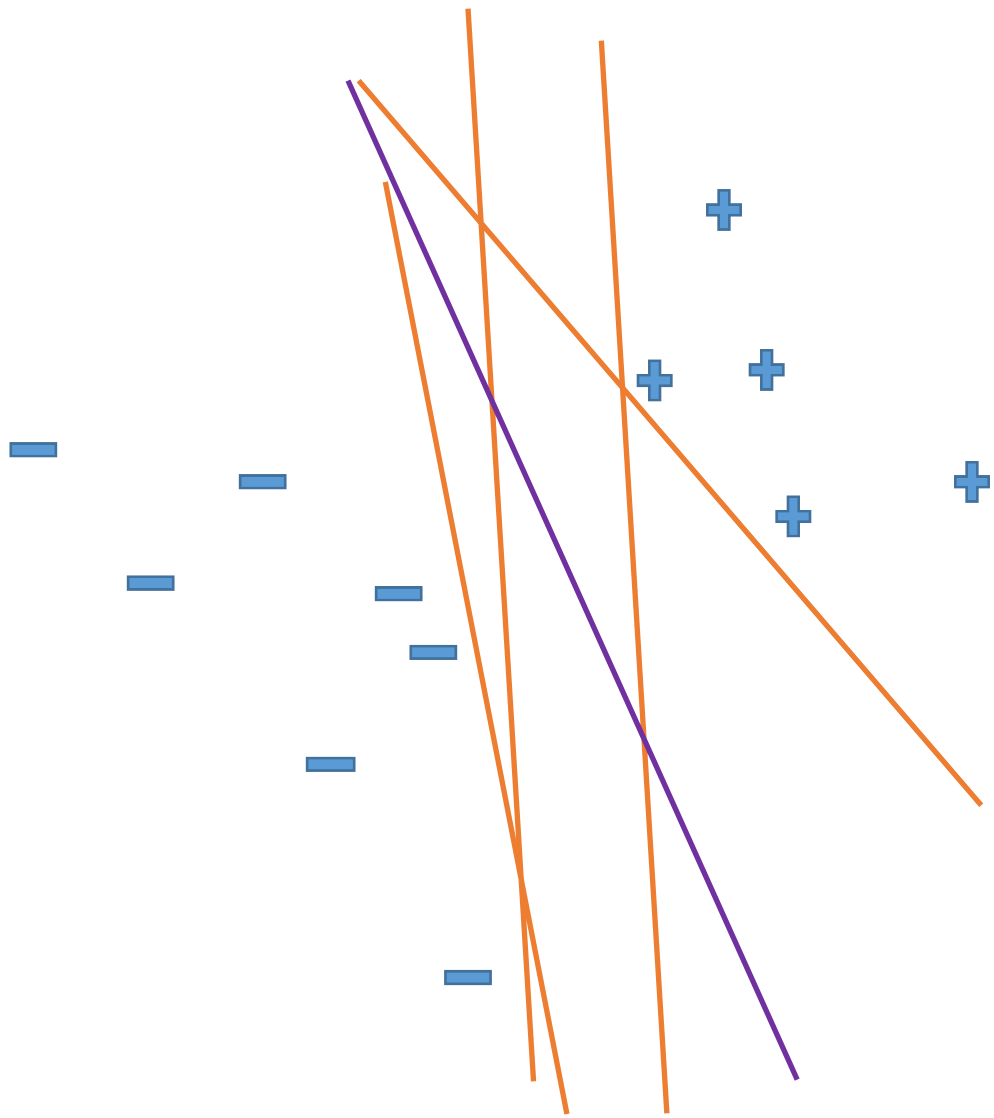
Assume data linearly separable

What is the best linear classifier?

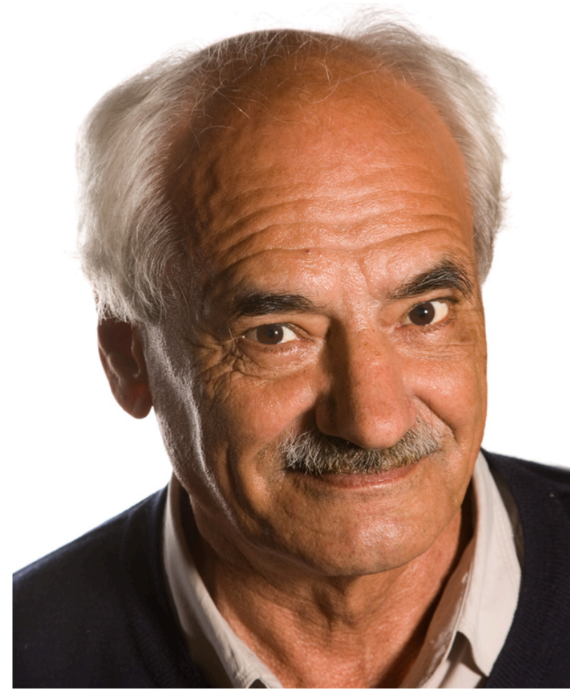


Assume data linearly separable

What is the best linear classifier?



Vapnik & Chervonenkis ‘63



Vapnik & Cortes

MM Hyper plane

Of all the hyper planes that separate the data, pick the one that has the highest margin.

Natural, but there is more.

Showed that choosing this hyperplane gives **good** generalization (small test error) under some statistical setting.

Developed much more … (will see soon)

How to find a max-margin plane?

Take a hyper plane $(\overline{w}, t)$ in $\mathbb{R}^d$.

Distance of $\bar{X} \in \mathbb{R}^d$ from this hyper plane is:

$$\frac{|\overline{w} \cdot \bar{X} - t|}{\| \overline{w} \|}$$

How to find a max-margin plane?

If $(\overline{w}, t)$ separates S, then for all $i = 1, \ldots, n$:

$$y_i = \mathrm{sign}(\overline{w} \cdot \bar{X}_i - t)$$

Distance of the point $\bar{X}_i$:

$$\gamma_i = \frac{y_i(\overline{w} \cdot \bar{X}_i - t)}{\| \overline{w} \|}$$

Margin: $\min \gamma_i$

How to find a max-margin plane?

Distance of the point $\bar{X}_i$:

$$\gamma_i = \frac{y_i(\bar{w} \cdot \bar{X}_i - t)}{\| \bar{w} \|}$$

Margin: $\geq \gamma$, when for all i:

$$\frac{y_i(\bar{w} \cdot \bar{X}_i - t)}{\| \bar{w} \|} \geq \gamma$$

How to find a max-margin plane?

Maximize over $\overline{w}, t, \gamma$:

$$\gamma$$

Subject to:

$$\frac{y_i(\overline{w} \cdot \bar{X}_i - t)}{\| \overline{w} \|} \geq \gamma$$

How to find a max-margin plane?

Maximize over $\vec{w}, t, \gamma$:

$$\gamma$$

Subject to:

$$y_i(\overline{w} \cdot \bar{X}_i - t) \geq \gamma$$

$$\| \overline{w} \| = 1$$

This is BAD optimization problem (why?)

How to find a max-margin plane?

Maximize over $\vec{w}, t, \gamma$:

$$\gamma$$

Subject to:

$$\frac{y_i(\bar{w} \cdot \bar{X}_i - t)}{\| \bar{w} \|} \geq \gamma$$

How to find a max-margin plane?

Maximize over $\overline{w}, t, \gamma$:

$$\frac{\gamma}{\| \overline{w} \|}$$

Subject to:

$$y_i(\overline{w} \cdot \bar{X}_i - t) \geq \gamma$$

How to find a max-margin plane?

We now normalize:

Maximize over $\overline{w}, t$:

$$\frac{1}{\| \overline{w} \|}$$

Subject to:

$$y_i(\overline{w} \cdot \bar{X}_i - t) \geq 1$$

How to find a max-margin plane?

We now normalize:

Minimize over $\overline{w}, t$:

$$\| \overline{w} \|_2^2$$

Subject to:

$$y_i(\overline{w} \cdot \bar{X}_i - t) \geq 1$$

This is GOOD optimization problem (why?)

Convex optimization