Outline

- Maximizing the margin

- The solution

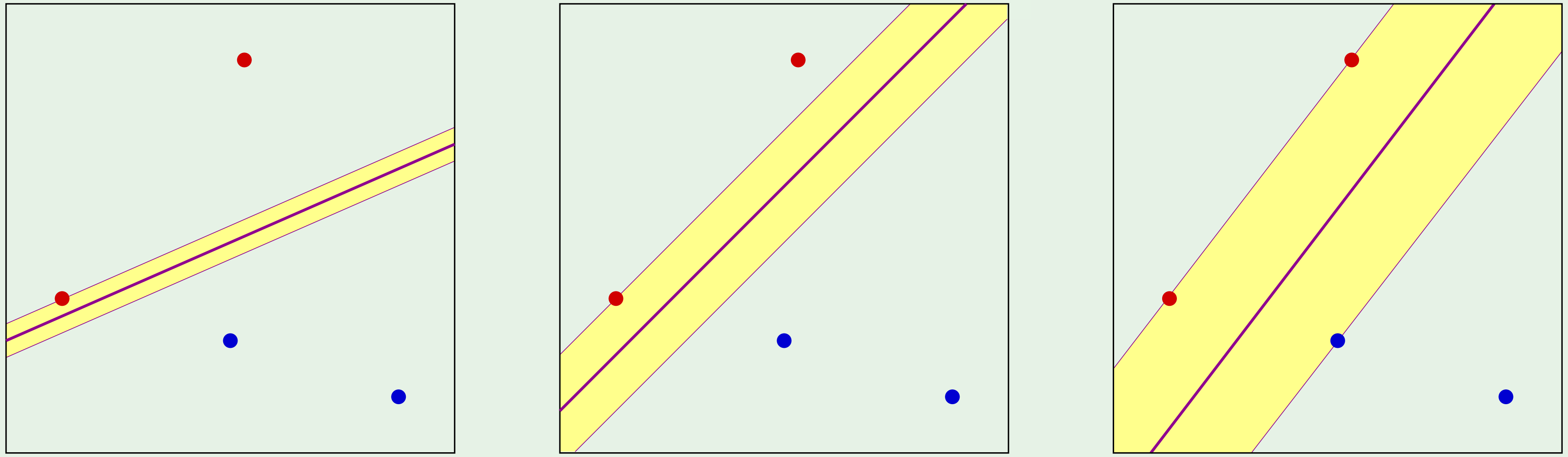
- Nonlinear transforms

Better linear separation

Linearly separable data

Different separating lines

Which is best?



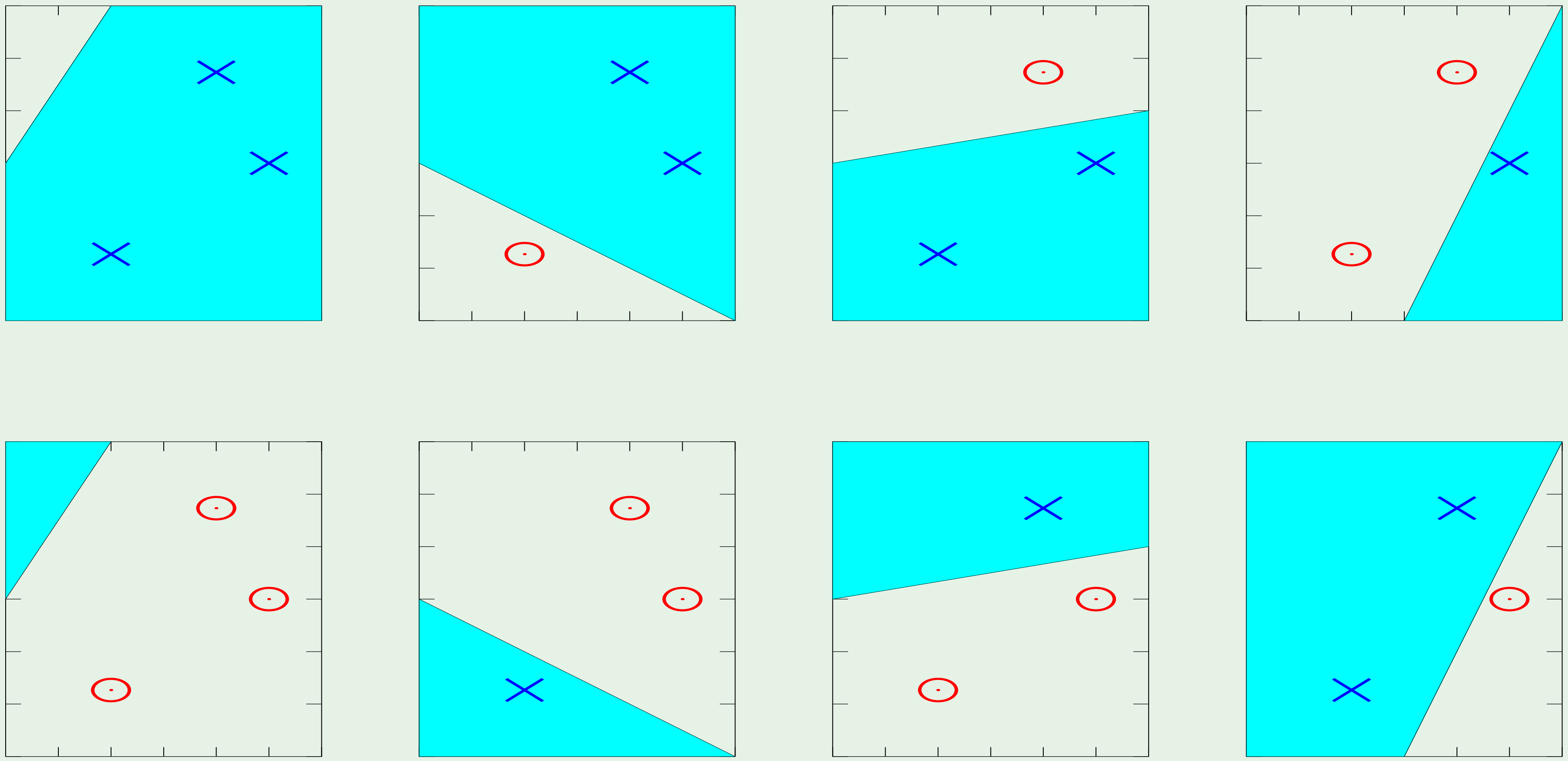
Two questions:

1. Why is bigger margin better?

2. Which $\mathbf{w}$ maximizes the margin?

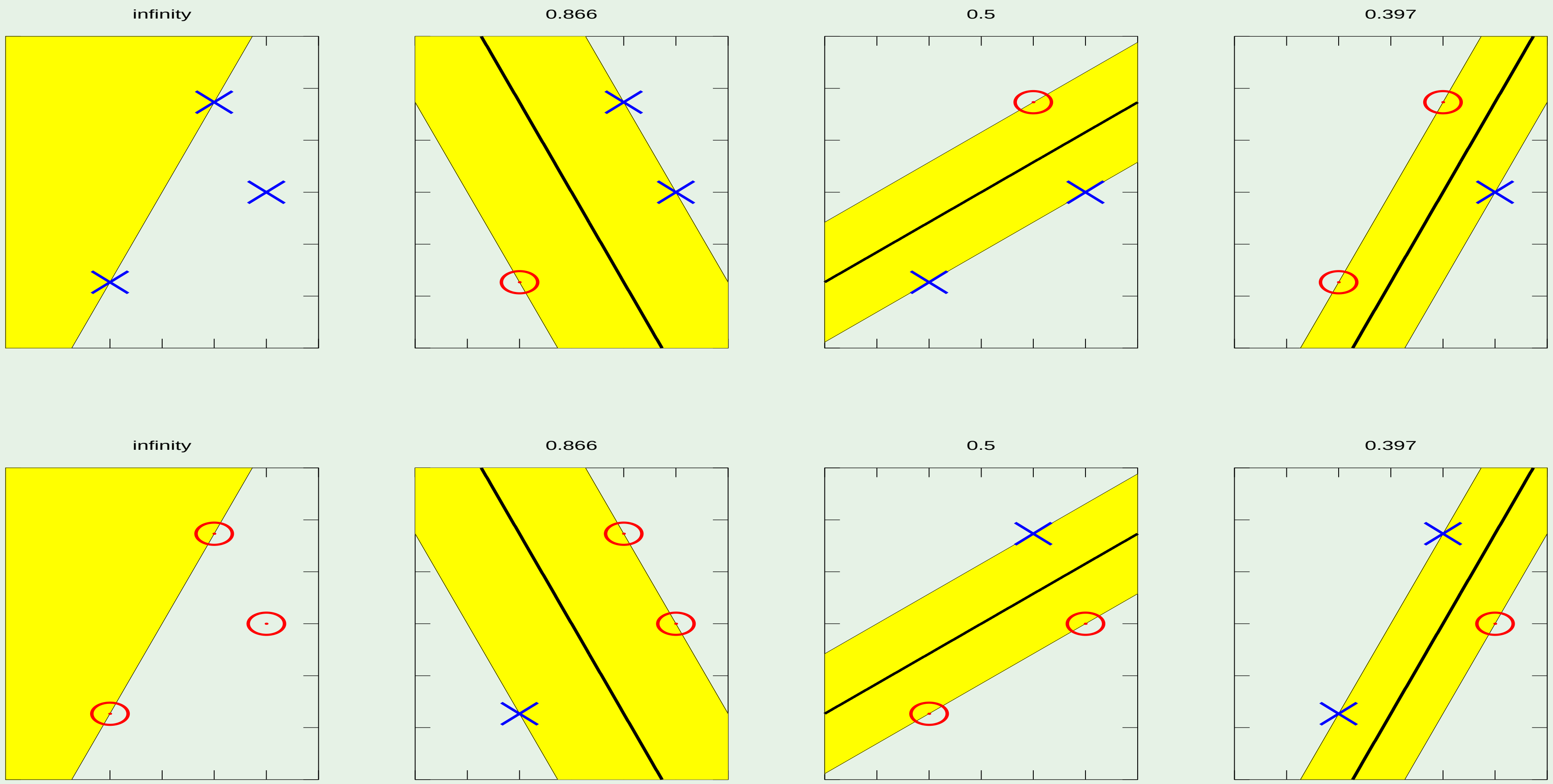
Remember the growth function?

All dichotomies with any line:



Dichotomies with fat margin

Fat margins imply fewer dichotomies



Finding $\mathbf{w}$ with large margin

Let $\mathbf{x}_n$ be the nearest data point to the plane $\mathbf{w}^\intercal\mathbf{x} = 0$. How far is it?

2 preliminary technicalities:

1. **Normalize $\mathbf{w}$**:

   $$|\mathbf{w}^\intercal\mathbf{x}_n| = 1$$

2. **Pull out w_0**:

   $\mathbf{w} = (w_1, \cdots, w_d)$ apart from b

   The plane is now $\boxed{\mathbf{w}^\intercal\mathbf{x} + b = 0}$ (no x_0)

Computing the distance

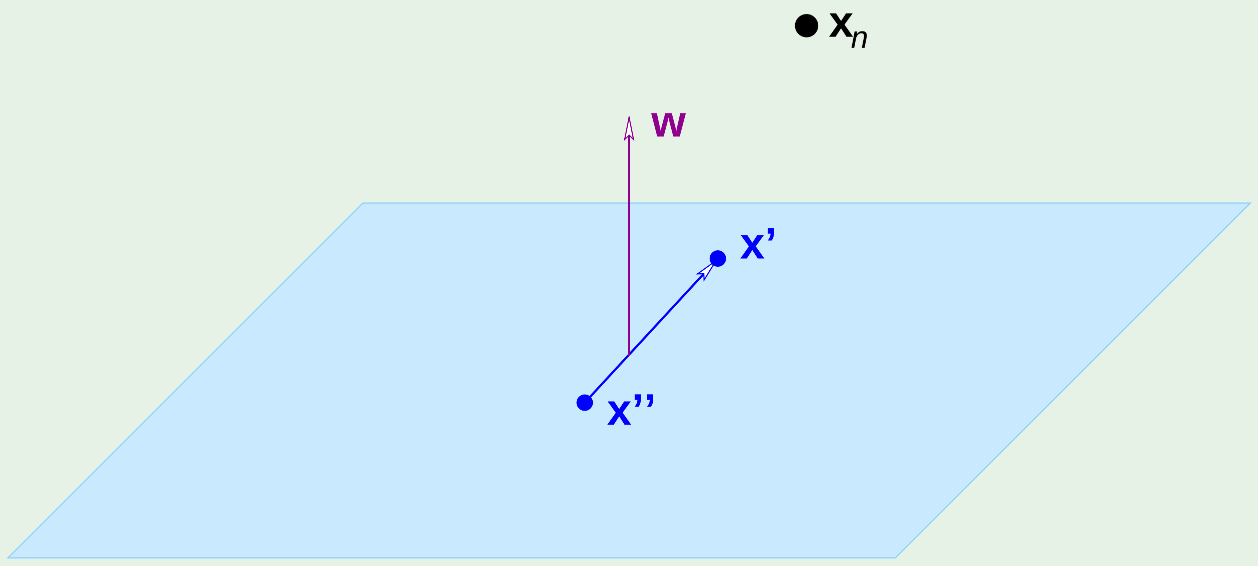
The distance between $\mathbf{x}_n$ and the plane $\mathbf{w}^\intercal\mathbf{x} + b = 0$ where $|\mathbf{w}^\intercal\mathbf{x}_n + b| = 1$

The vector $\mathbf{w}$ is $\perp$ to the plane in the $\mathcal{X}$ space:

Take $\mathbf{x}'$ and $\mathbf{x}''$ on the plane

$$\mathbf{w}^\intercal\mathbf{x}' + b = 0 \quad \text{and} \quad \mathbf{w}^\intercal\mathbf{x}'' + b = 0$$

$$\Longrightarrow \quad \mathbf{w}^\intercal(\mathbf{x}' - \mathbf{x}'') = 0$$

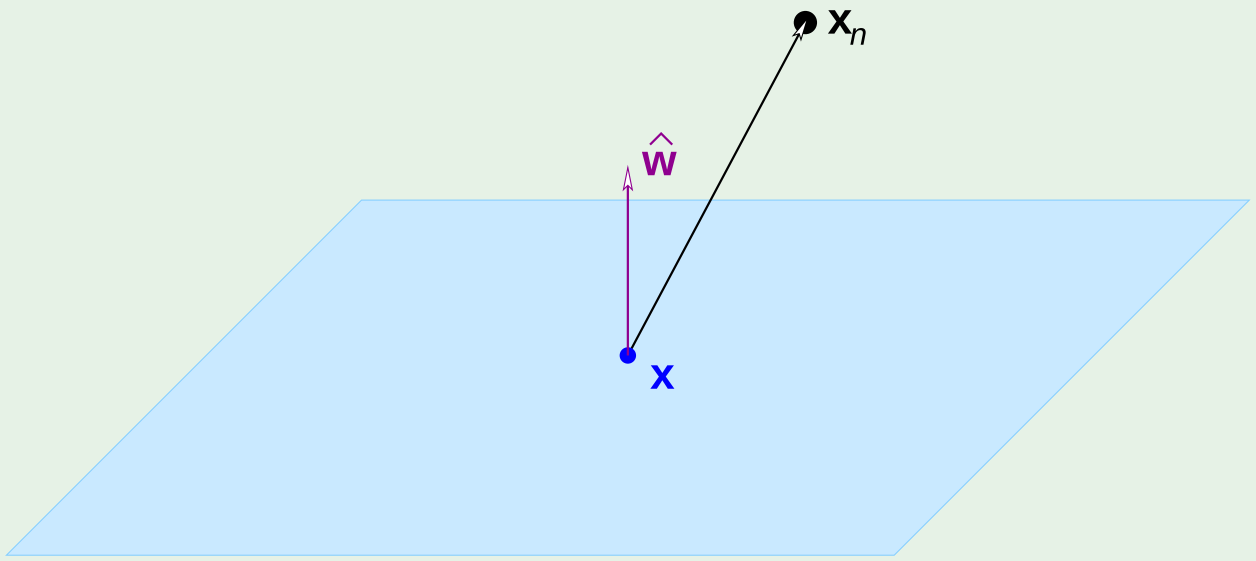


and the distance is ...

Distance between $\mathbf{x}_n$ and the plane:

Take any point $\mathbf{x}$ on the plane

Projection of $\mathbf{x}_n - \mathbf{x}$ on $\mathbf{w}$



$$\hat{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|} \implies \text{distance} = \left|\hat{\mathbf{w}}^\intercal(\mathbf{x}_n - \mathbf{x})\right|$$

$$\text{distance} = \frac{1}{\|\mathbf{w}\|}\left|\mathbf{w}^\intercal\mathbf{x}_n - \mathbf{w}^\intercal\mathbf{x}\right| = \frac{1}{\|\mathbf{w}\|}\left|\mathbf{w}^\intercal\mathbf{x}_n + b - \mathbf{w}^\intercal\mathbf{x} - b\right| = \frac{1}{\|\mathbf{w}\|}$$

The optimization problem

$$\text{Maximize } \frac{1}{\|\mathbf{w}\|}$$

$$\text{subject to } \min_{n=1,2,\ldots,N} \left|\mathbf{w}^\intercal \mathbf{x}_n + b\right| = 1$$

Notice: $\left|\mathbf{w}^\intercal \mathbf{x}_n + b\right| = y_n \left(\mathbf{w}^\intercal \mathbf{x}_n + b\right)$

$$\text{Minimize } \frac{1}{2} \mathbf{w}^\intercal \mathbf{w}$$

$$\text{subject to } y_n \left(\mathbf{w}^\intercal \mathbf{x}_n + b\right) \geq 1 \quad \text{for} \quad n = 1, 2, \ldots, N$$