How can SVM maximize the margin in decision boundary by minimizing the $\sum_{j=1}^{d} \theta_j$?

One thing to notice that this minimization is attempted with the constraints of:

- $y_i = 1 \implies \theta^{\mathrm{T}} x_i \geq 1$ ("predictions of positive examples should give a value that ≥ 1") and
- $y_i = -1 \implies \theta^{\mathrm{T}} x_i \leq -1$ ("predictions of negative examples should give a value that ≤ -1").

This means that, the value of prediction should provide sufficient **quantity** for each training example. This quantity $\theta^{\mathrm{T}} x_i$, geometrically, is the **projection of x_i onto the "parameter vector" θ**
.

The length of a projection is determined by:

- the length of participating vectors,
- the angle they form.

Now that the length of one participating vector, x_i, is fixed (i.e. determined by input training data), to make the projection meet the requirement for quantity, we can either:

- make sure θ is big, or
- make sure the angle is right.

The first idea is not graceful: it basically looks like as if the model is standing in front of a crowd of audiences and shouting out "YES YES YES! THIS x_i WORKS!!" Instead, we want the angle to be better positioned. Thus, we seek to minimize θ.