CME 216, ME 343 - Winter 2020

Eric Darve, ICME
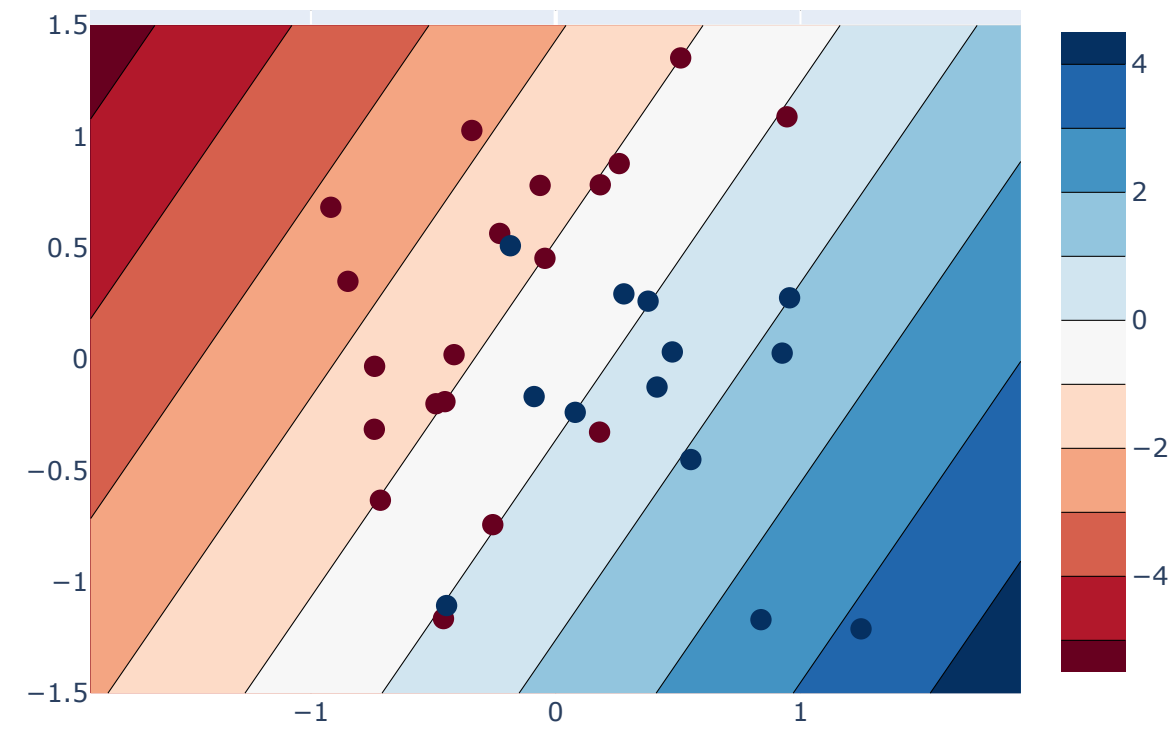
Soft-margin

For many problems, though, because of noise and complex data, it is not possible to have a hyperplane that exactly separates the data.

In that case, there is no solution to the optimization problem above.

Decision function

The figure shows an example where no line divides the red dots from the blue dots.

The optimization problem from the previous section has no solution in that case.

One solution is to introduce slack variables so that some constraints can be violated but in a minimal way

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

with $\xi_i \geq 0$.

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

If the constraint $y_i(w^T x_i + b) \geq 1$ can be satisfied, then $\xi_i = 0$.

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

If $1 > \xi_i > 0$, then the constraint is weakly violated but the classification is still correct.

The sign of $w^T x_i + b$ (which is used to predict the label) is still the same as y_i.

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

But, if $\xi_i > 1$ then the data is misclassified.

The sign of $w^T x_i + b$ is now different from the sign of y_i.

Hopefully this only happens for very few points.

The new optimization problem becomes:

$$(w, b, \xi) = \mathrm{argmin}_{w,b,\xi} \frac{1}{2}|w|_2^2 + C\sum_{i=1}^{n}\xi_i$$

with the following constraints:

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

C is set by the user and determines how much slack we are allowing.

A large C means little violation is tolerated.

Very few points are allowed to violate the condition.

The hyperplane is strongly determined by the points nearest to the hyperplane.

The hyperplane is very sensitive to points that violate the condition.

A small C means a lot of violations are possible.

Small C is required when data has a lot of noise that needs to be filtered out.

In that case, many violations will be accepted as long as this leads to a large separation $1/|w|_2$.

With noise
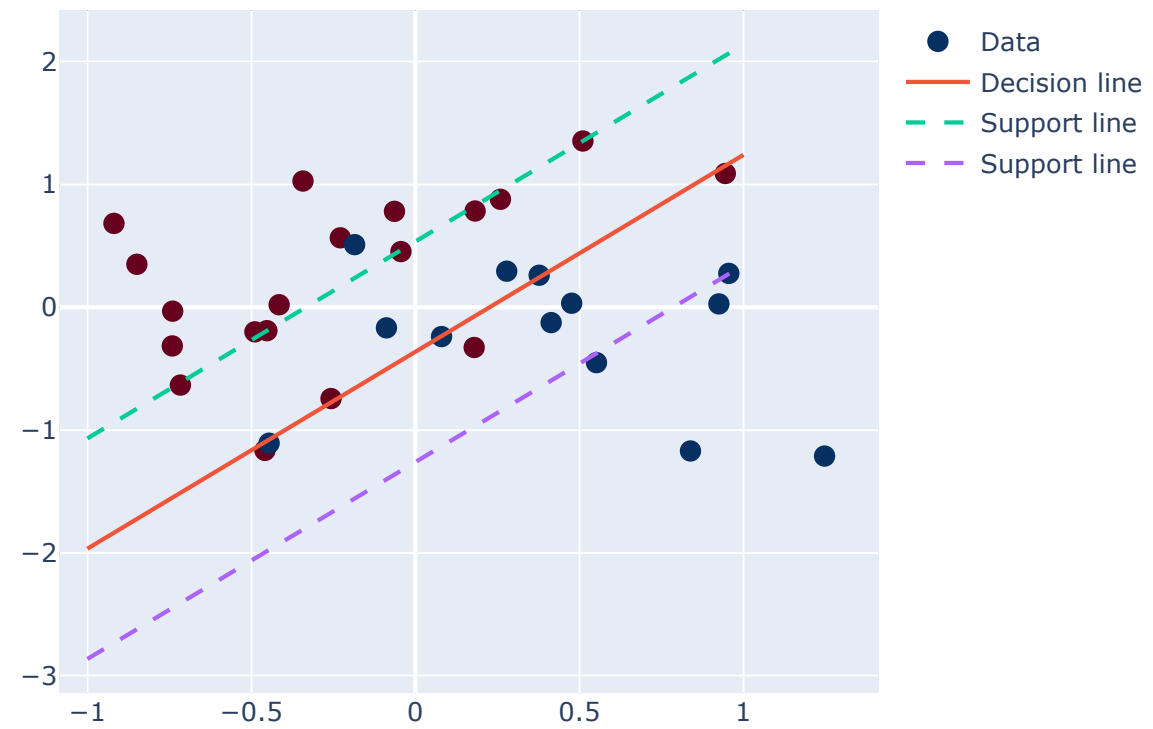
Now we see that points are allowed to lie between the green and purple lines.

There are even a few red points below the orange line and a few blue points above.

But this cannot be avoided since no line perfectly separates the red points from the blue points.