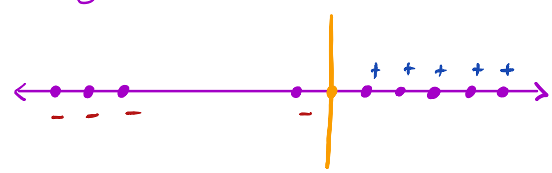
# Soft Margin SVM

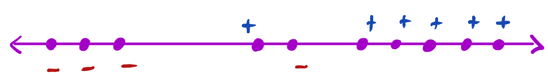
Suponga el siguiente dataset con una feature:



Aquí el punto naranja sería la frontera de decisión, que genera la división sin error de mayor margen. ¿Es este un modelo que generaliza bien?

Probablemente no. De hecho este es un ejemplo de bajo Bias, pero alta varianza. Veamos otro problema del SVM.

¿Cómo podemos separar este dataset?



En efecto, no podemos hacer una separación como las vistas hasta ahora: no hay forma de maximizar el margen con 0 error.

Para obtener un modelo que generalice mejor y pueda separar el segundo ejemplo vamos a permitir pequeños errores.

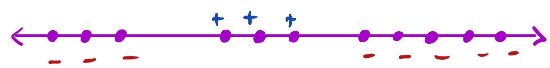
Así, un Soft Margin SVM es un modelo más general que balancea entre generar una calle "lo más ancha posible" pero que busca tolerar pocos elementos dentro de la calle. Usualmente, queremos pocos elementos "en la calle" (en inglés, "Margin Violations") pero esto nos puede llevar a generalizar peor. Esto lo controla un parámetro C que viene del problema de optimización:

$$\min_{\vec{\omega}, b, \xi} \frac{1}{2}\|\vec{\omega}^T\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.a.} \quad y_i(\vec{\omega}^T x_i + b) \geq 1 - \xi_i \qquad \xi_i \geq 0 \qquad \forall i = 1, \dots, n$$

Ahora en el problema de optimización introducimos la variable $\xi_i$ que mide cuanto la instancia $i$ tiene permitido salirse del margen. Si C es muy grande, esos errores van a pesar más, por lo que tendremos potencialmente un margen menor, pero que quizás haga que el modelo no generalice bien.

Ahora, ¿qué pasa en este caso?



Un clasificador lineal no podrá hacer un buen trabajo. Pero veremos un truco que se puede aplicar al modelo SVM para resolver esta situación.