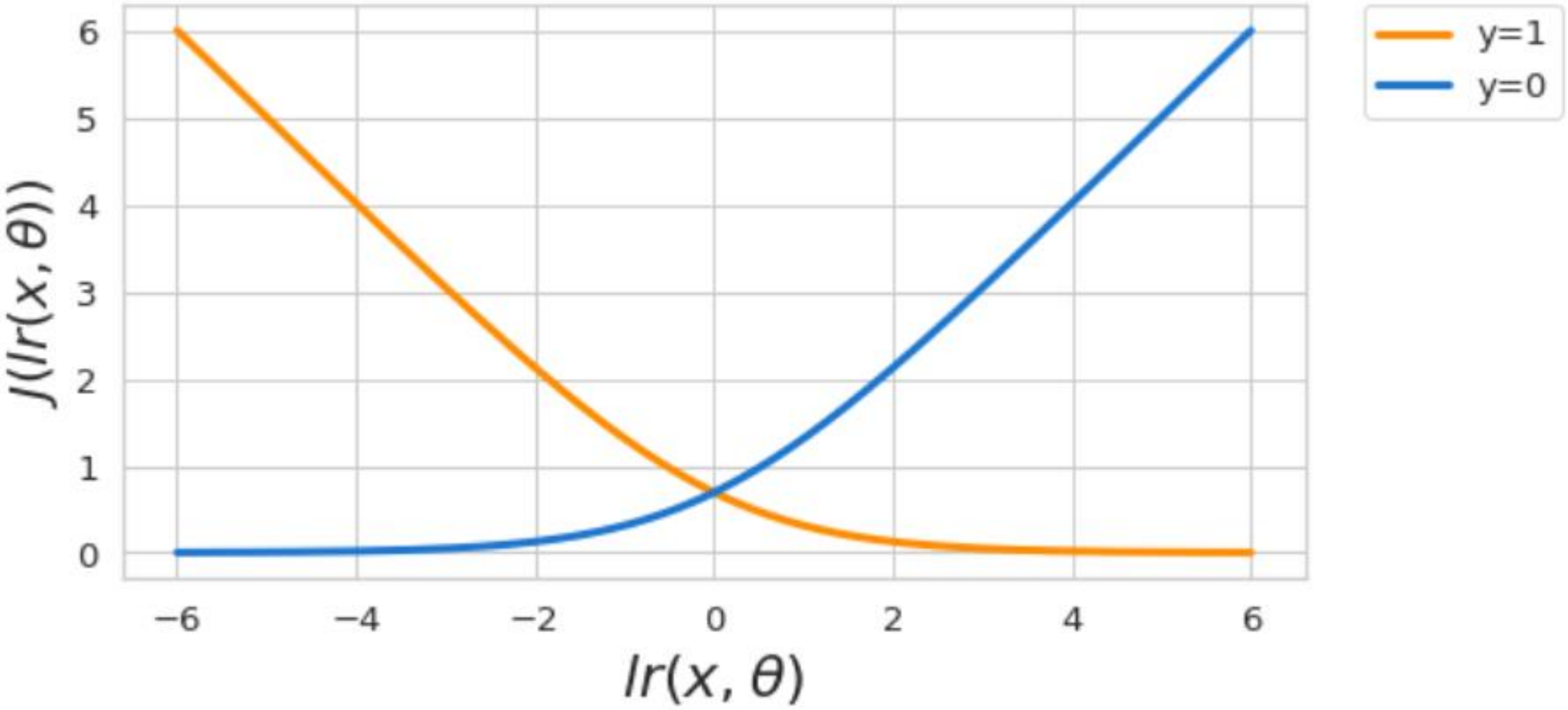
logistic regression: cost (loss function)

$$J(\theta) = -[\frac{1}{n}\sum_{i=1}^{n} y^{(i)} log(f(x^{(i)}, \theta)) + (1 - y^{(i)}) log(1 - f(x^{(i)}, \theta))]$$

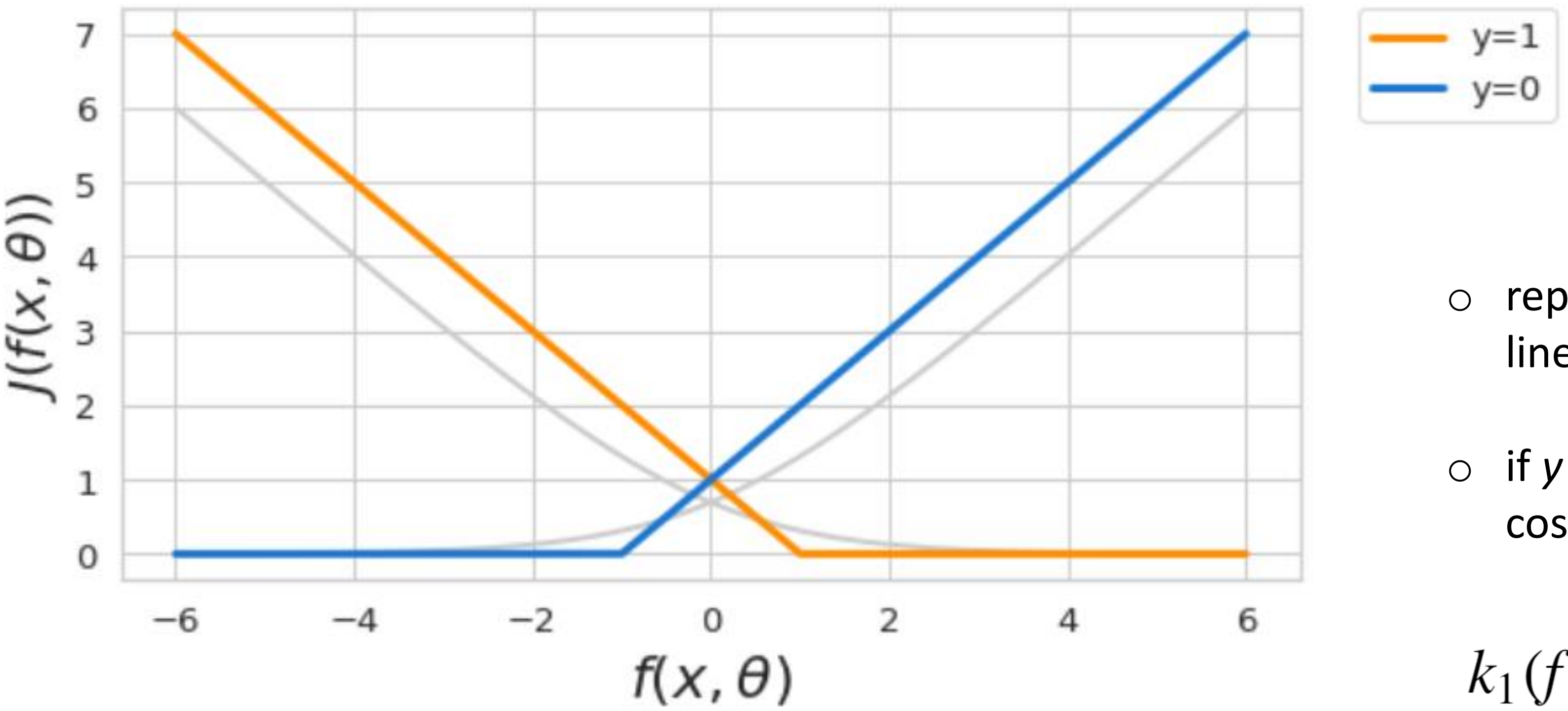
We know that $y^{(i)}$ is either 0 or 1. If $y^{(i)} = 1$ then the cost function $J(\theta)$ is incremented by $-log(f(x^{(i)}, \theta))$.

Similarly, if $y^{(i)} = 0$ then the cost function $J(\theta)$ is incremented by $-log(1 - f(x^{(i)}, \theta))$.

logistic regression: cost (loss function)



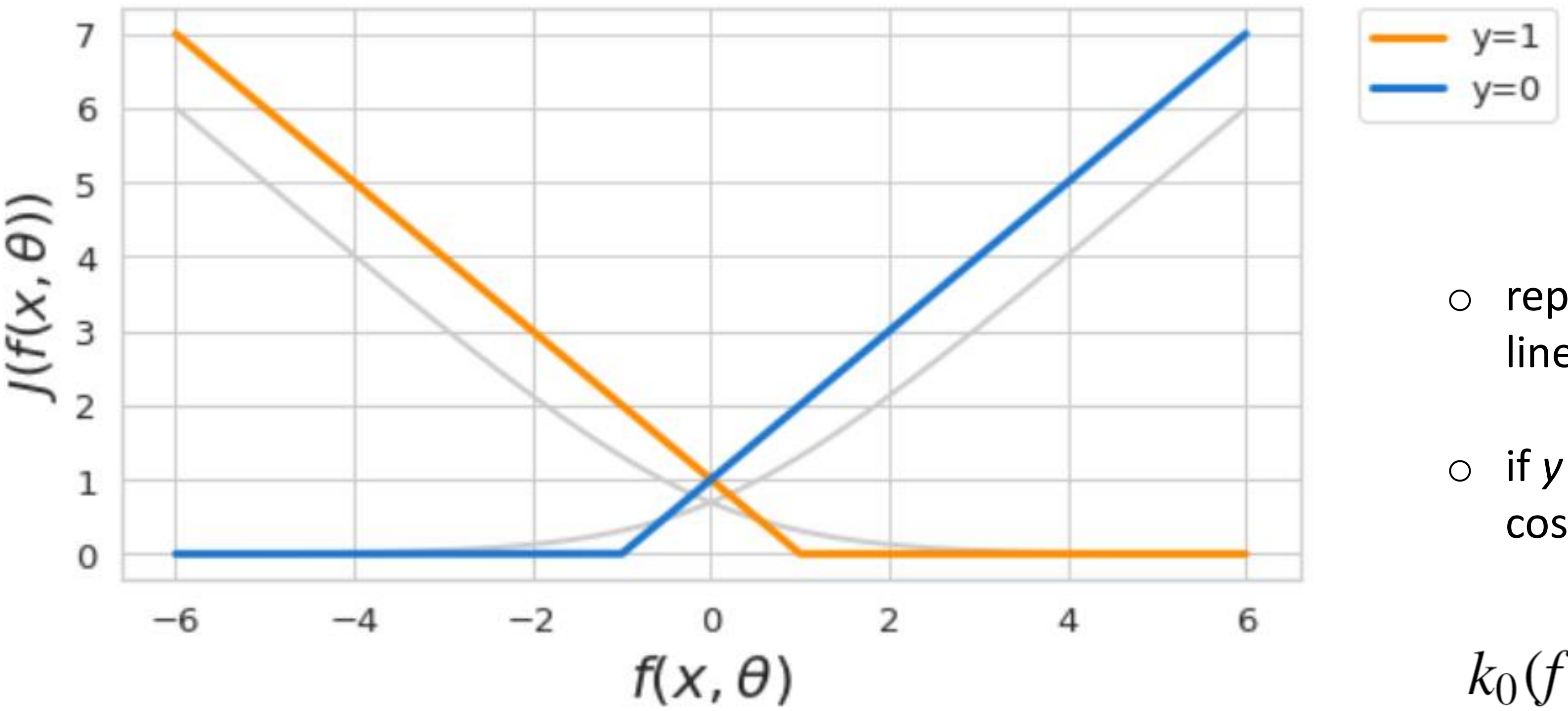
Support vector machine: cost (loss function)



- replace cost function by piecewise linear function
- if y = 1 then the contribution to the cost is

$$k_1(f(x,\theta)) = max(0, 1 - f(x,\theta))$$

Support vector machine: cost (loss function)



- replace cost function by piecewise linear function
- if $y = 0$ then the contribution to the cost is

$$k_0(f(x,\theta)) = max(0, 1 + f(x,\theta))$$

Support vector machine: cost (loss function)

Fit a linear model

$$f(x, \theta) = \theta' x$$

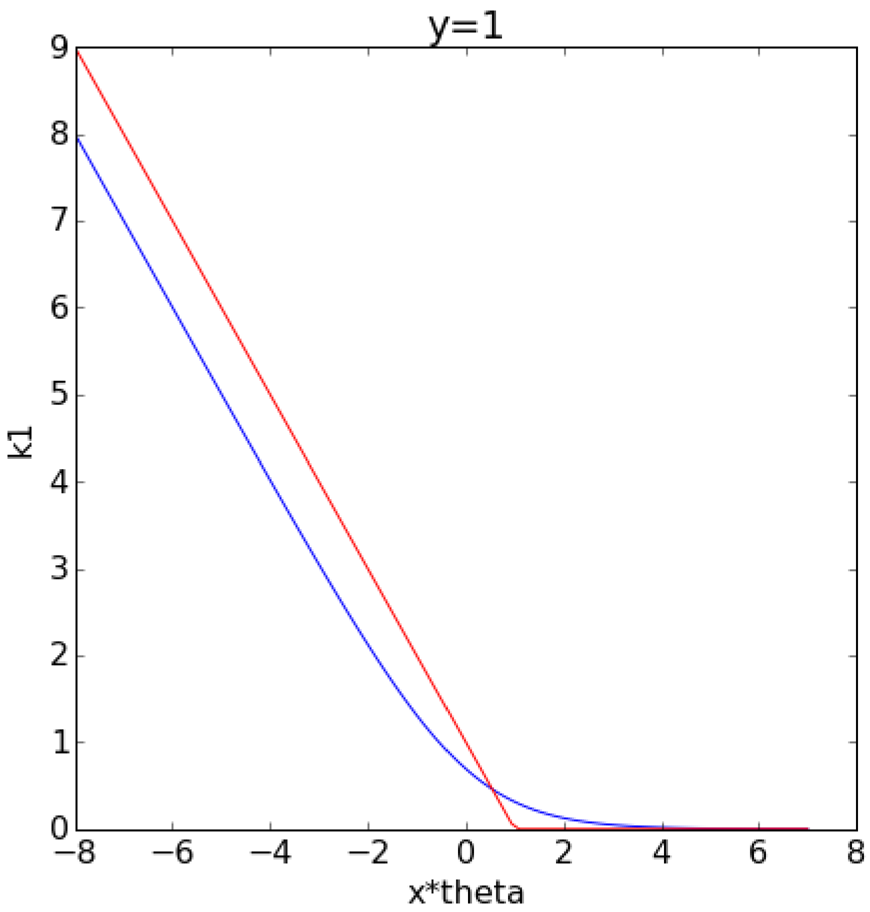
such that

$$J(\theta) = \left[C \sum_{i=1}^{n} y^{(i)} k_1(\theta' x^{(i)}) + (1 - y^{(i)}) k_0(\theta' x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^{m} \theta_j^2$$

with $k_1(\theta' x) = max(0, 1 - \theta' x)$ and $k_0(\theta' x) = max(0, 1 + \theta' x)$

is minimized.

Support vector machine: cost (loss function)

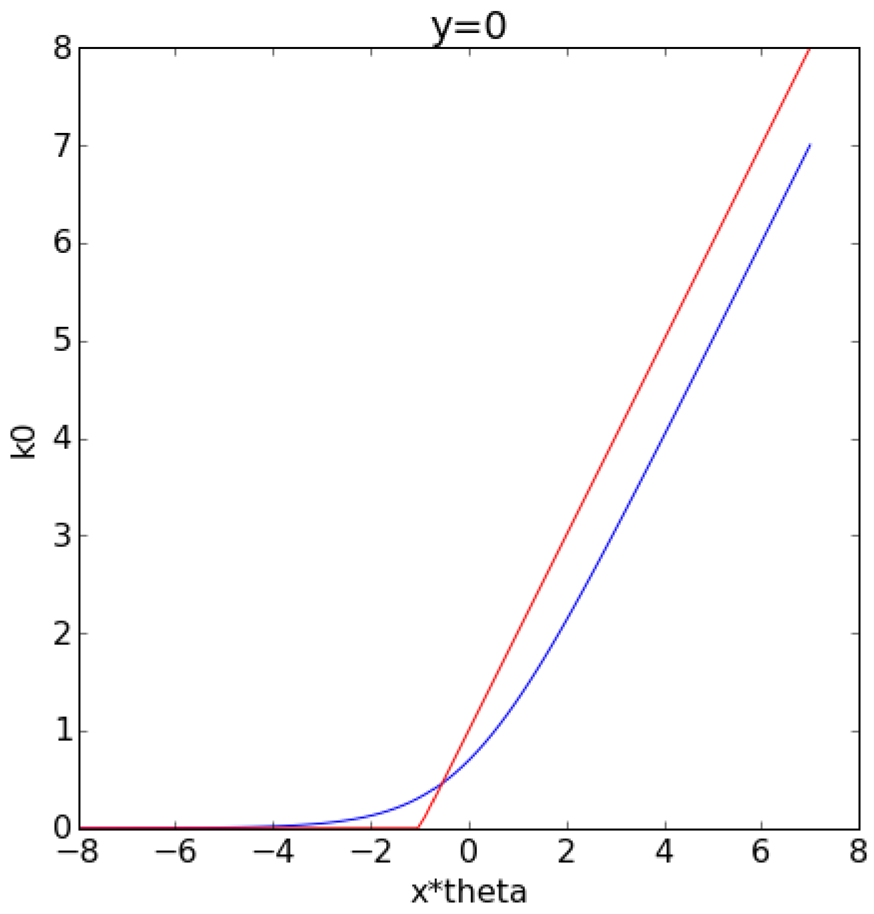


- In this case the contribution to the cost needs to be small when the model predicts high values (>0) and large when the model predicts low values (<0).

- For SVMs we see that the contribution to cost decreases linearly and becomes zero when

$$\theta' x \geq 1$$

Support vector machine: cost (loss function)



- In this case the contribution to the cost needs to be large when the model predicts high values (>0) and small when the model predicts low values (<0).

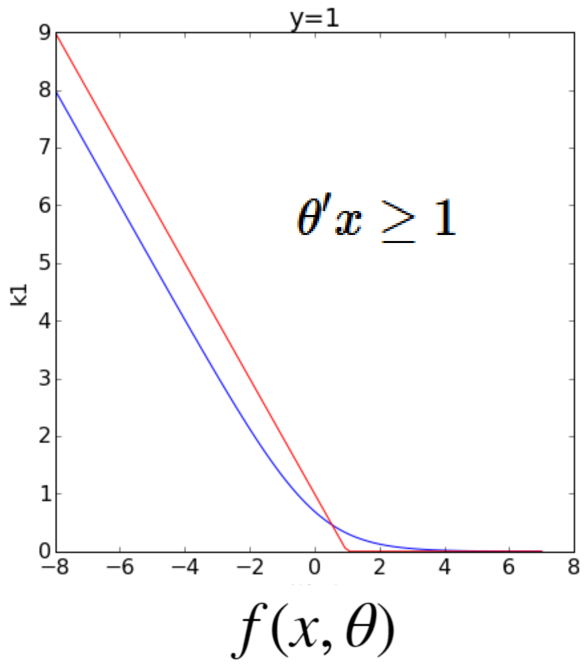
- For SVMs we see that the contribution to the cost is zero for

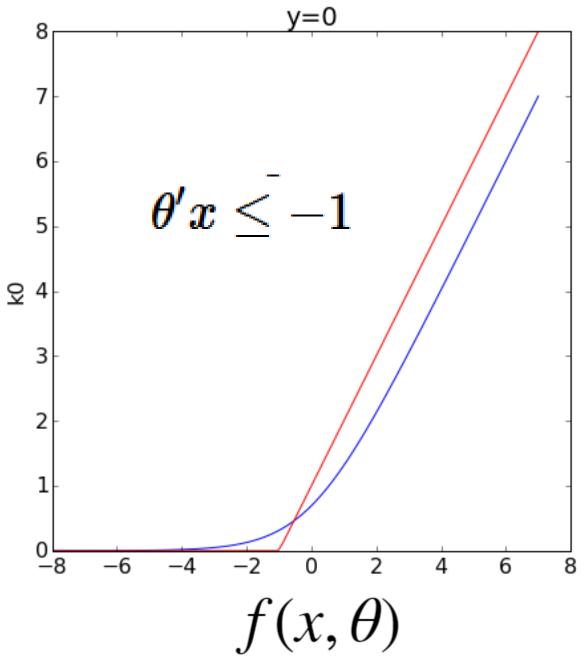
$$\theta' x \leq -1$$

and then increases linearly.

Support vector machine: cost (loss function)

$$J(\theta) = \left[C \sum_{i=1}^{n} y^{(i)} k_1(\theta' x^{(i)}) + (1 - y^{(i)}) k_0(\theta' x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^{m} \theta_j^2$$

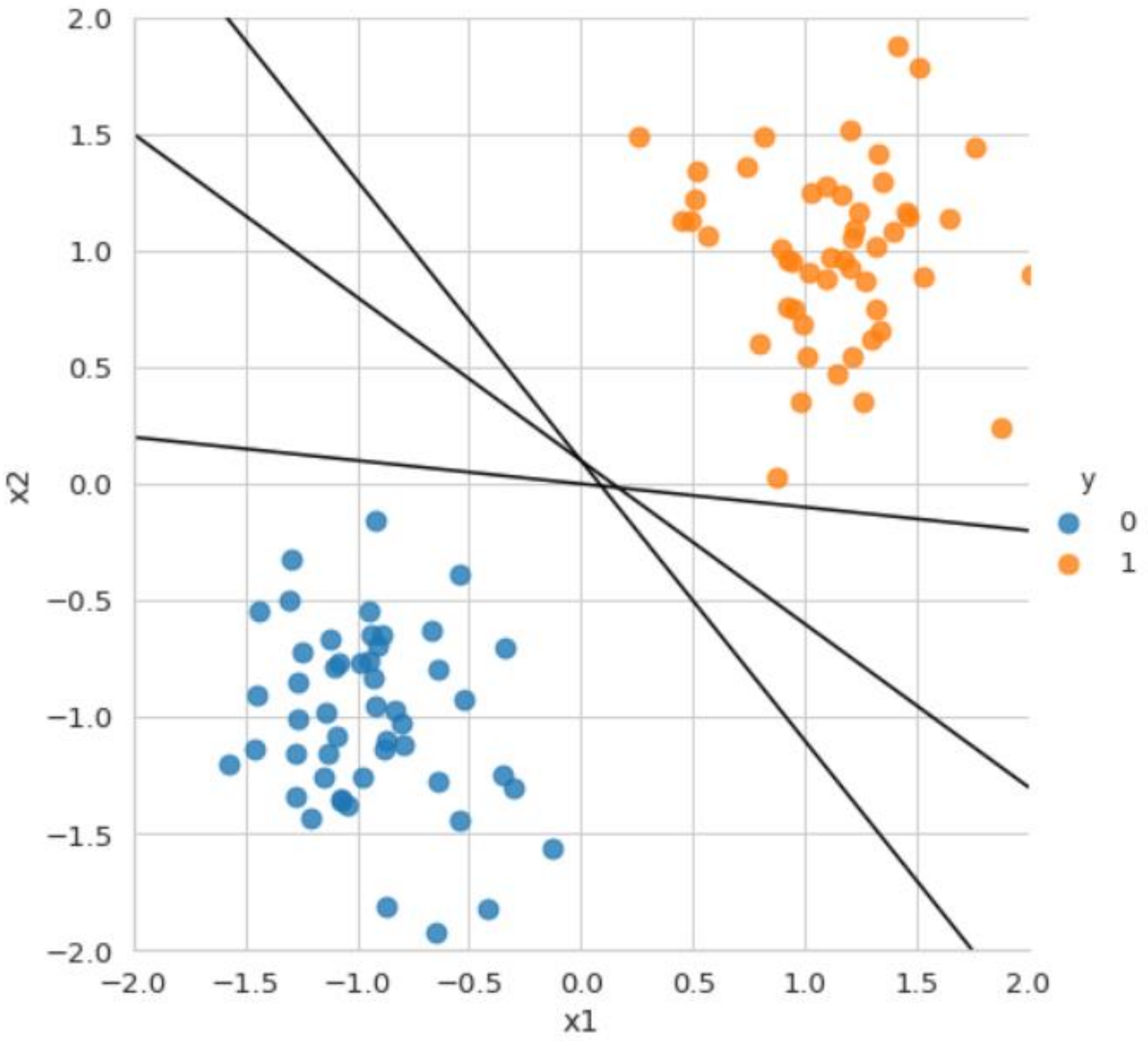


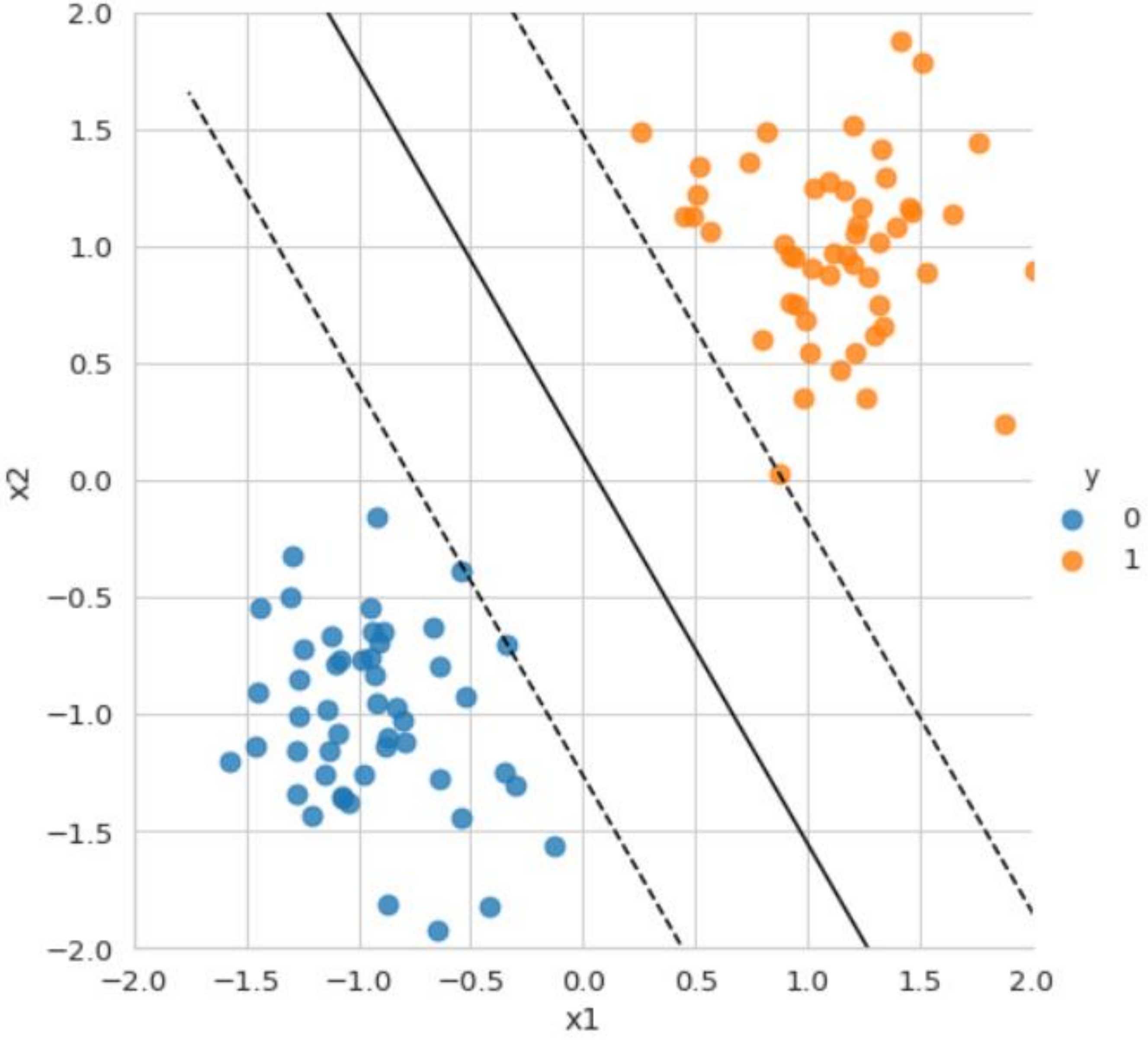


- Consider a train set with two classes that are perfectly linearly separable.
- The piecewise cost function can be made zero.
- The SVM objective can be written as

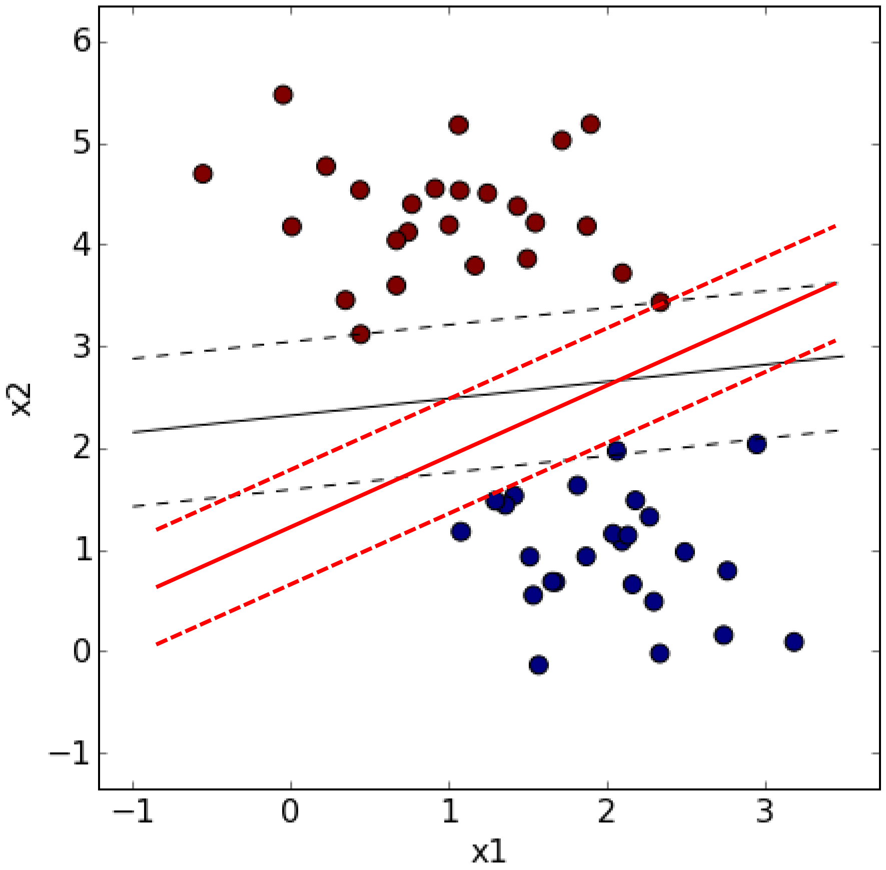
$$\min_{\theta} \frac{1}{2} \sum_{j=1}^{m} \theta_j^2$$

$$\text{subject to } \theta' x^{(i)} \geq 1 \text{ if } y^{(i)} = 1$$
$$\theta' x^{(i)} \leq -1 \text{ if } y^{(i)} = 0$$

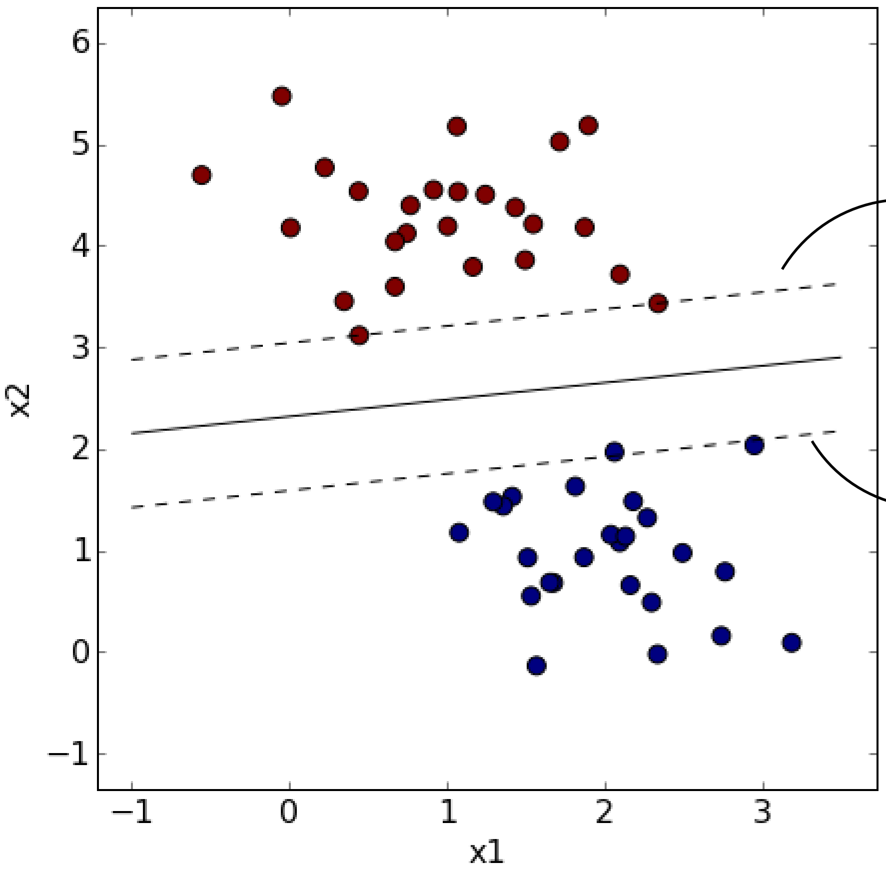

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
x2
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x1
y
0
1


2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
x2
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x1
y
0
1

Support vector machine: large margin classifier



Support vector machine: large margin classifier



$$\theta' x^{(i)} = 1$$

$$\theta' x^{(i)} = -1$$

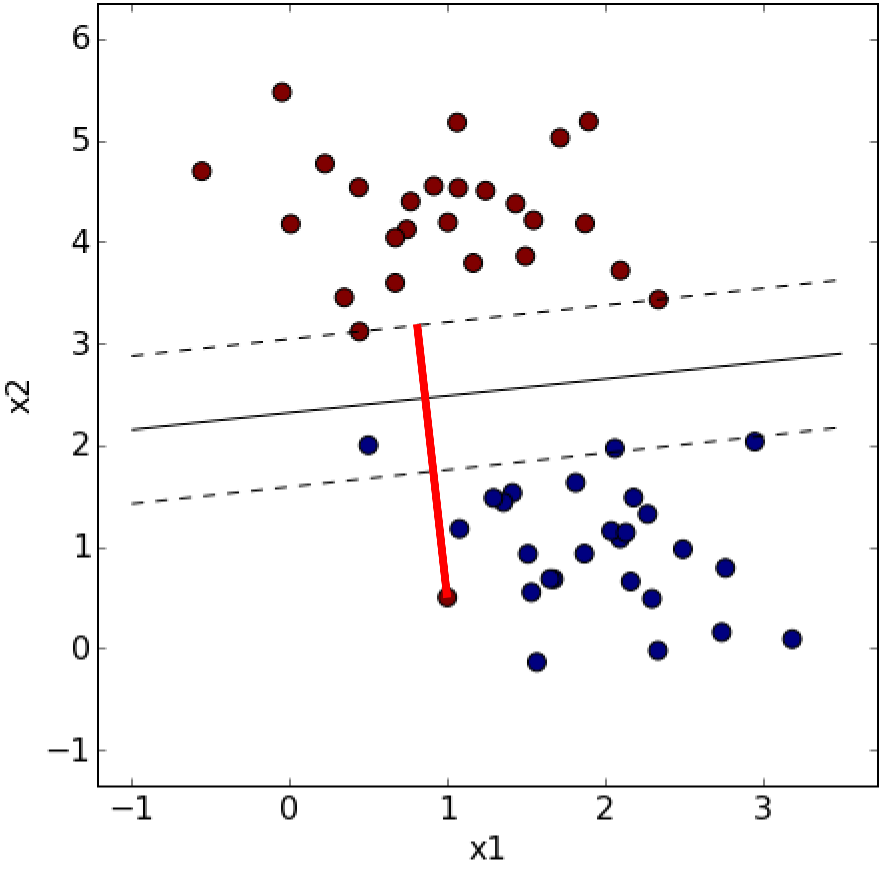
- red points statisfy $\theta' x^{(i)} \geq 1$
- blue points satisfy $\theta' x^{(i)} \leq -1$
- distance bewteen two dashed lines is

$$\frac{2}{\sqrt{\sum_{j=1}^{m} \theta_j^2}}$$

- SVM objective was

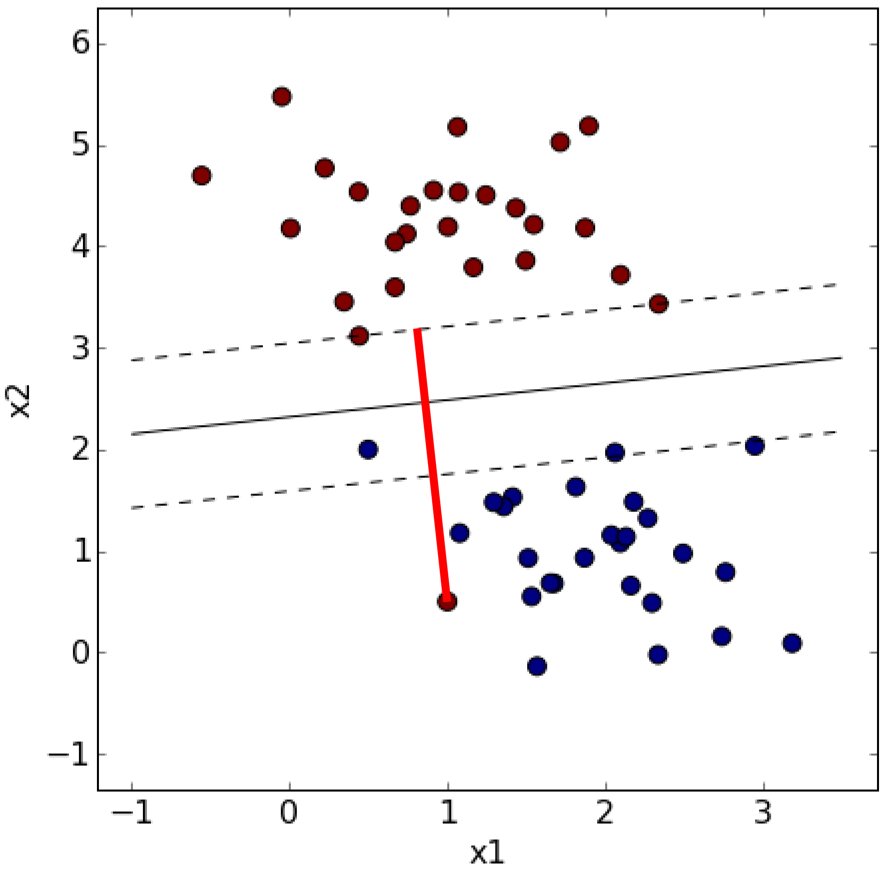
$$\min_{\theta} \frac{1}{2} \sum_{j=1}^{m} \theta_j^2$$

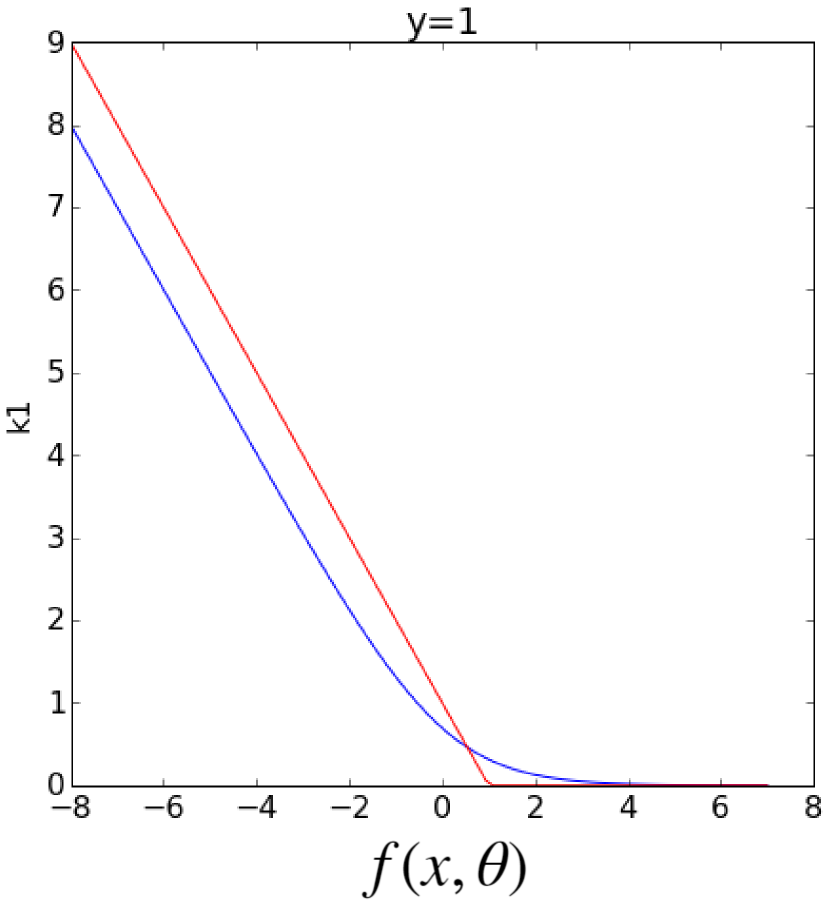
Support vector machine: soft margin



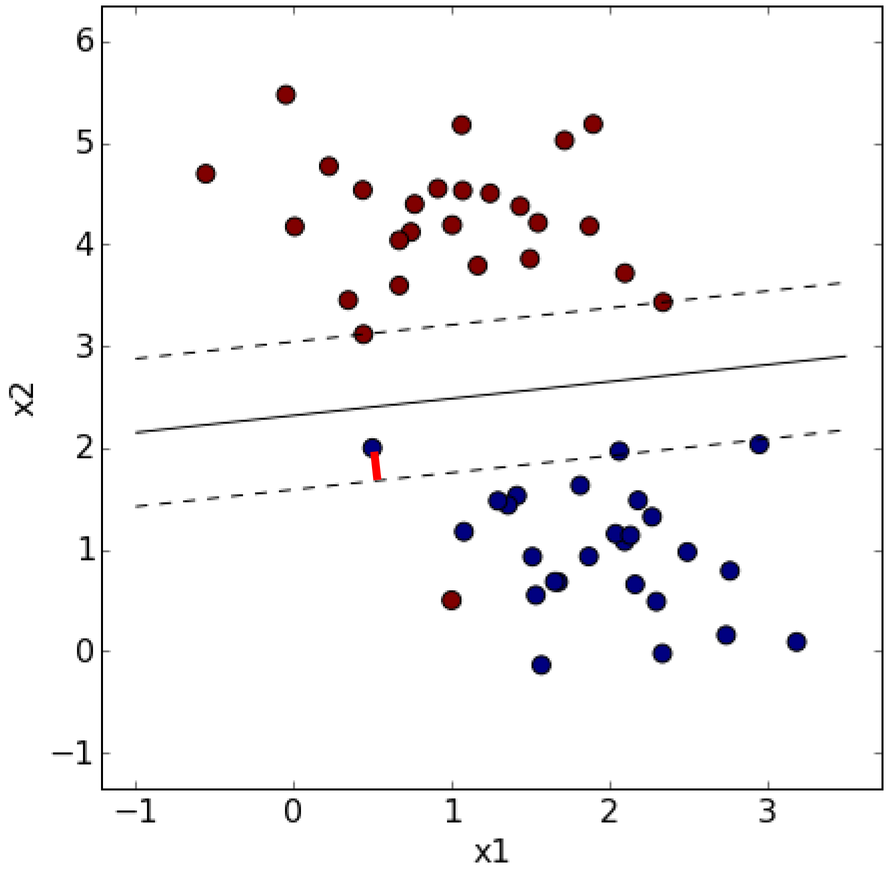
$$\theta' x^{(i)} \geq 1 \text{ if } y^{(i)} = 1$$
$$\theta' x^{(i)} \leq -1 \text{ if } y^{(i)} = 0$$

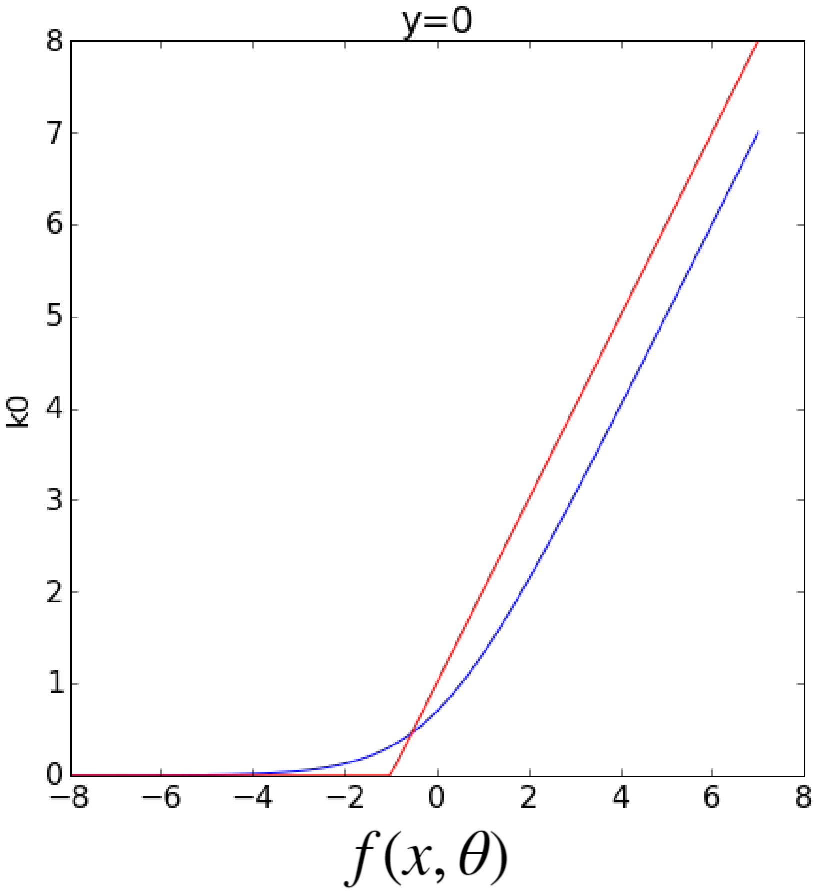
Support vector machine: soft margin



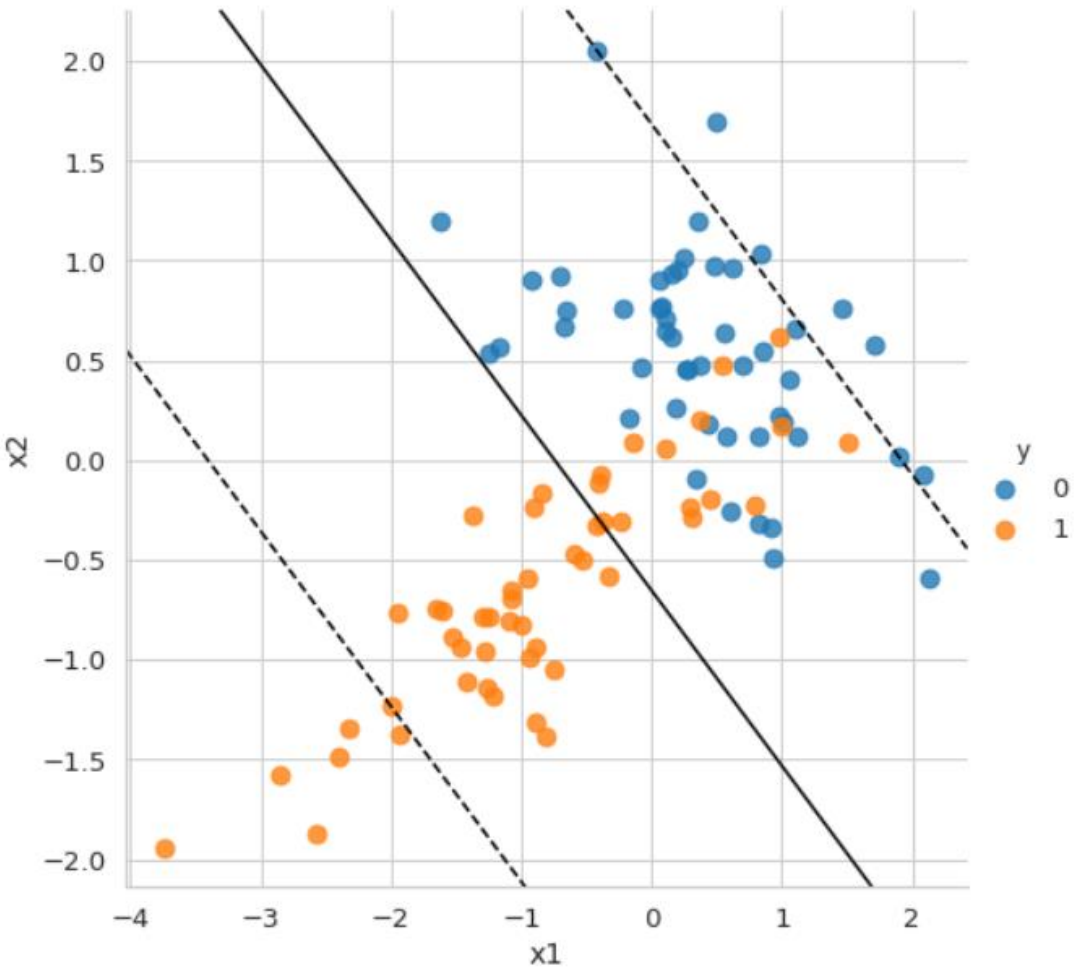


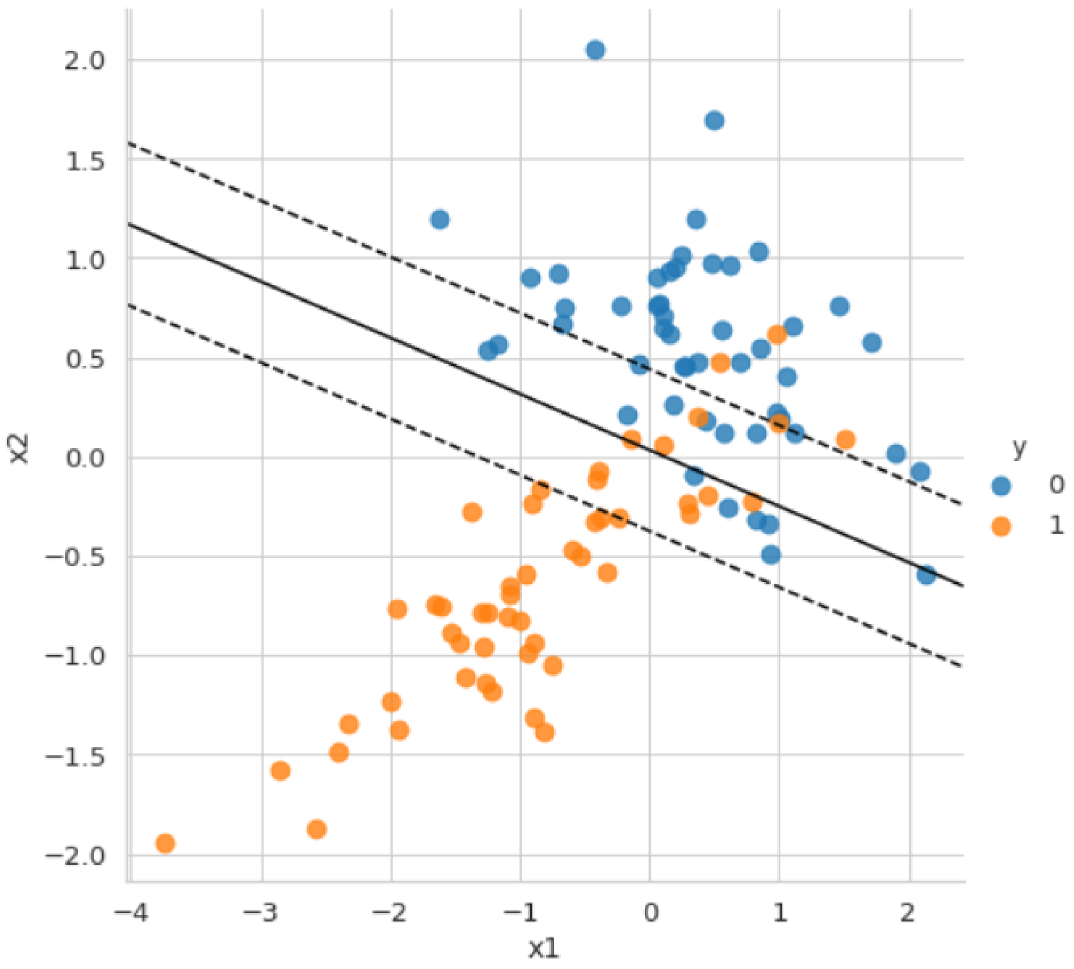
Support vector machine: soft margin



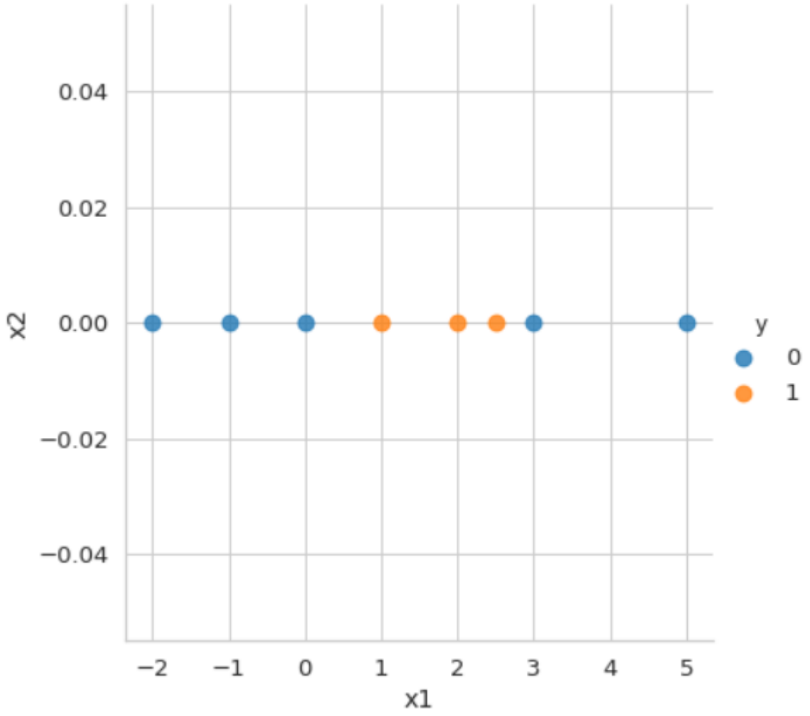


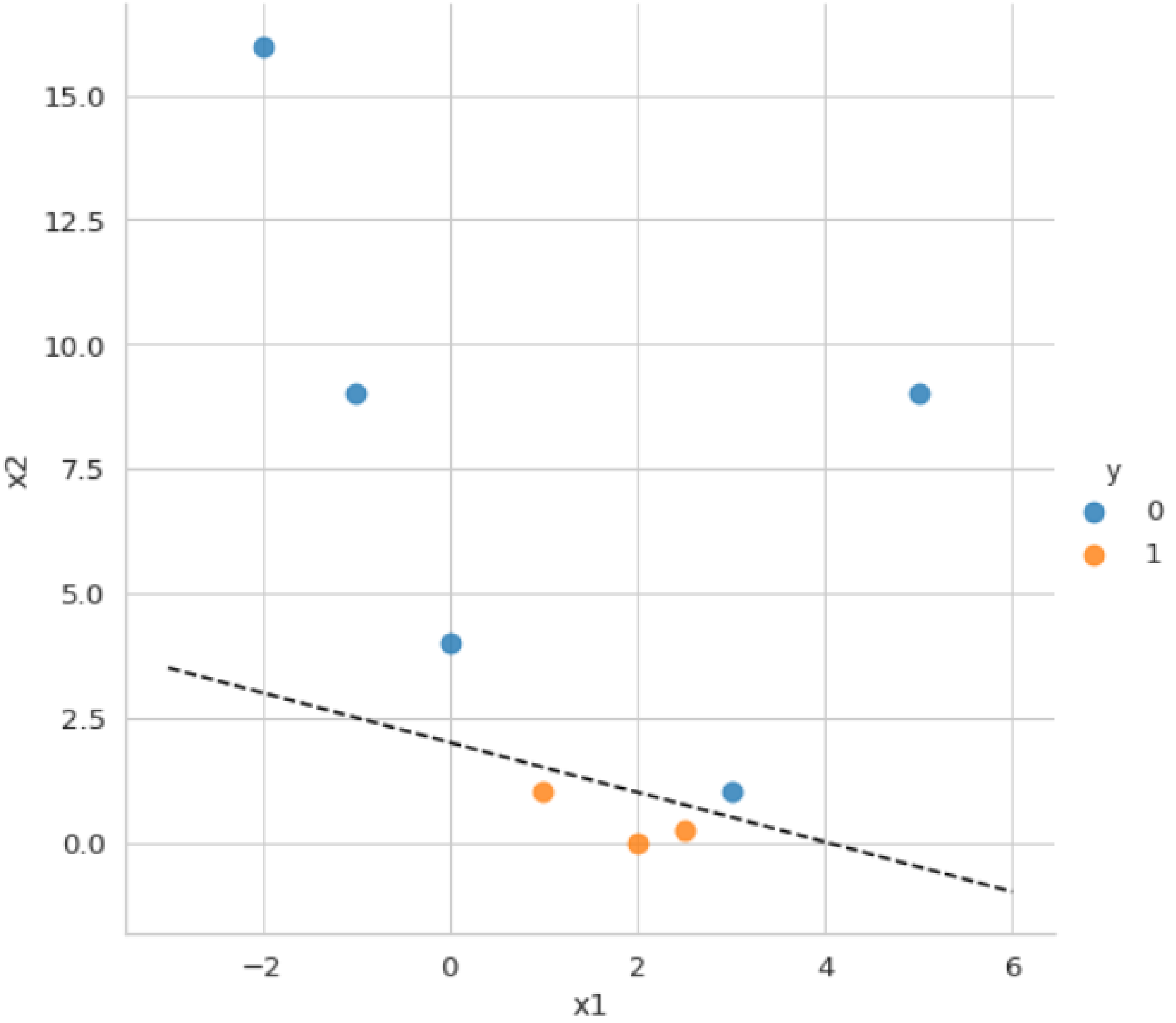
Support vector machine: regularization hyperparameter





Support vector machine: kernels





Support vector machine: kernels

SVMs can also be formulated as a linear function of the samples (dual form) instead of the features as

$$f(x,\theta) = \sum_{i=1}^{n} \theta_i (x \cdot x^{(i)}) + \theta_0$$

that can be reformulated as a non-linear function using what is known as a kernel function

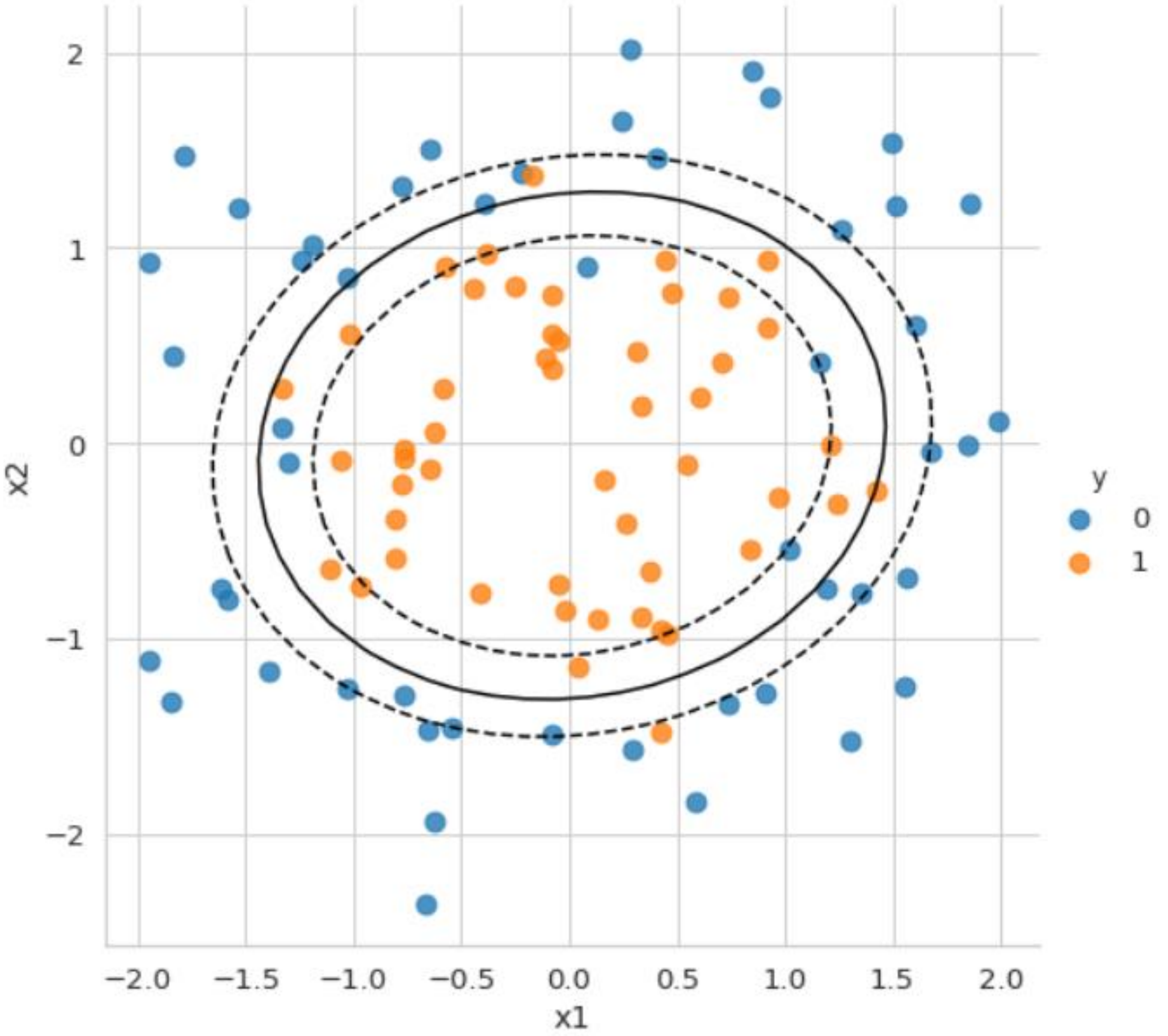
$$K(x^{(i)}, x^{(j)})$$

to become

$$f(x,\theta) = \sum_{i=1}^{n} \theta_i K(x, x^{(i)}) + \theta_0$$

The data points $x^{(i)}$ for which $\theta_i > 0$ are called the support vectors.

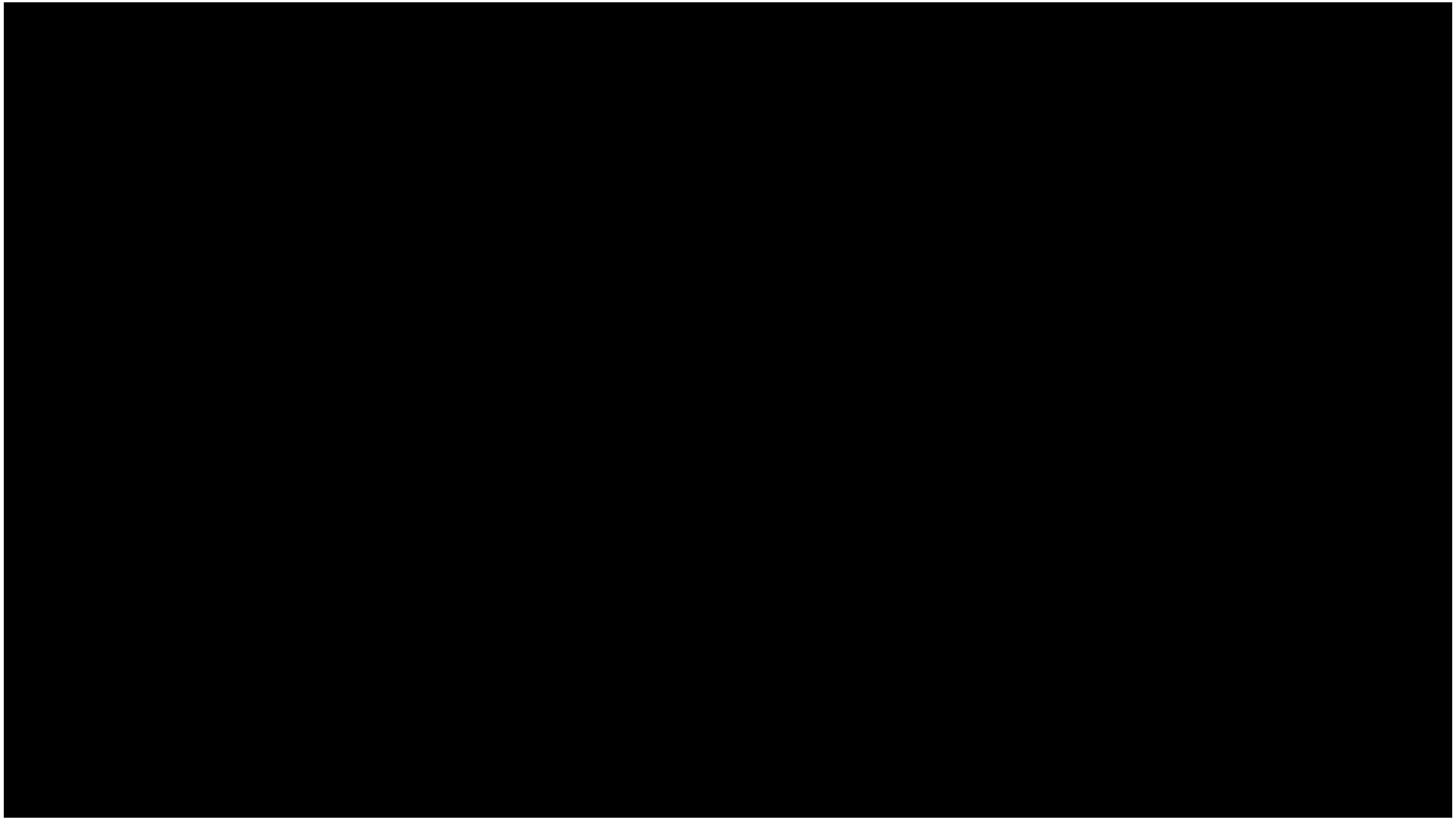
Support vector machine: polynomial kernel



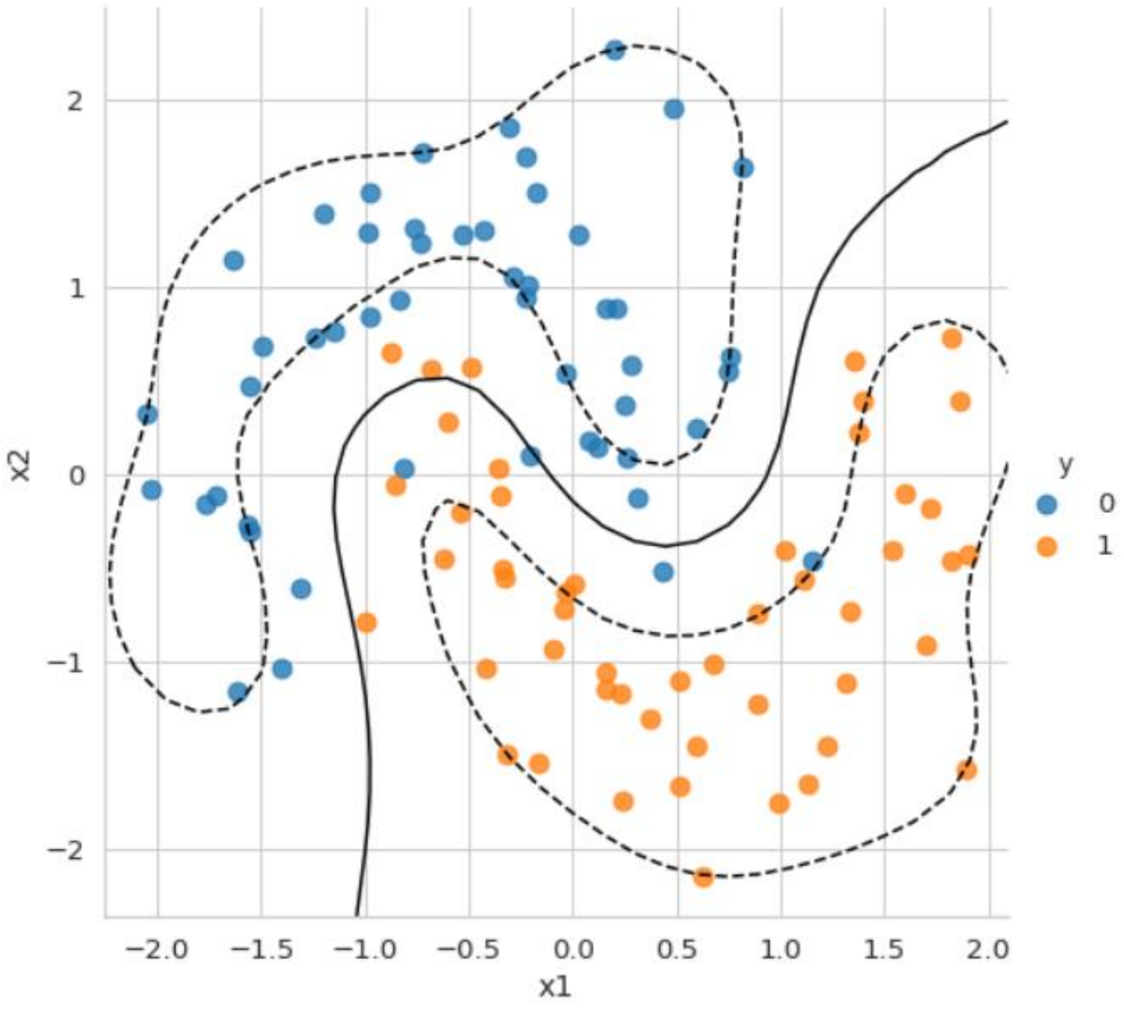
$$f(x,\theta) = \sum_{i=1}^{n} \theta_i K(x, x^{(i)}) + \theta_0$$

$$K(x^{(i)}, x^{(j)}) = (x^{(i)} \cdot x^{(j)} + c)^d$$

Support vector machine: polynomial kernel



Support vector machine: RBF kernel



$$f(x, \theta) = \sum_{i=1}^{n} \theta_i K(x, x^{(i)}) + \theta_0$$

$$K(x^{(i)}, x^{(j)}) = \exp\left[-\frac{||x^{(i)} - x^{(j)}||^2}{2\sigma^2}\right]$$

Support vector machine: kernels