Machine Learning Homework Sheet 07

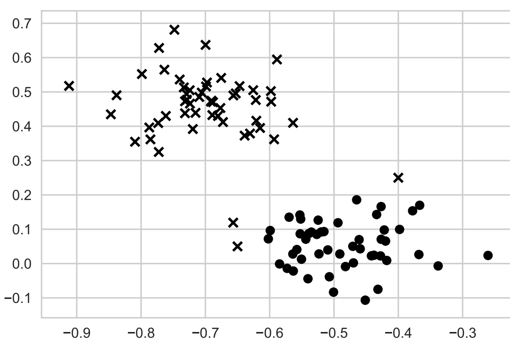
Soft-margin SVM and Kernels

1 Soft-margin SVM

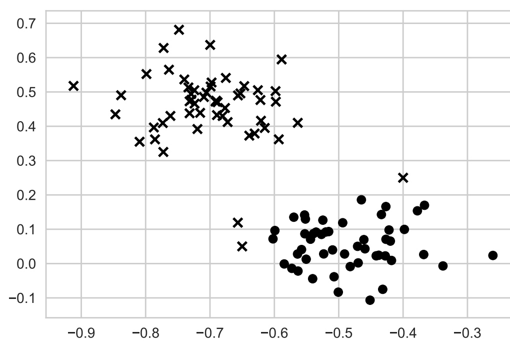
Problem 1: Assume that we have a linearly separable dataset $\mathcal{D}$, on which a soft-margin SVM is fitted. Is it guaranteed that all training samples in $\mathcal{D}$ will be assigned the correct label by the fitted model? Explain your answer.

Problem 2: Why do we need to ensure that $C > 0$ in the slack variable formulation of soft-margin SVM? What would happen if this was not the case?

Problem 3: Sketch the decision boundary of an SVM with a quadratic kernel (polynomial with degree 2) for the data in the figure below, for two specified values of the penalty parameter C. (The two classes are denoted as $\bullet$'s and $\times$'s.)



(a) $C = 10^{10}$



(b) $C = 10^{-10}$

Explain the reasoning behind your sketch of the decision boundary for both cases (one sentence for each plot).

2 Kernels

Problem 4: Show that for $N \in \mathbb{N}$ and $a_i \geq 0$, with $i \in [0, N]$ the function

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{i=1}^{N} a_i \left(\boldsymbol{x}_1^T \boldsymbol{x}_2\right)^i + a_0$$

is a valid kernel.

Problem 5: Find the feature transformation $\phi(x)$ corresponding to the kernel

$$k(x_1, x_2) = \frac{1}{1 - x_1 x_2},$$

with $x_1, x_2 \in (0, 1)$.

Hint: Consider an infinite-dimensional feature space.

Problem 6: Consider the following algorithm.

Algorithm 1: Counting something

input : Character string x of length m (one based indexing)
input : Character string y of length n (one based indexing)
output: A number $s \in \mathbb{R}$

```
s ← 0;
for i ← 1 to m do
    for j ← 1 to n do
        if x[i] == y[j] then
            s ← s + 1;
```

a) Explain, in no more than two sentences, what the above algorithm is doing.

b) Let $\mathcal{S}$ denote the set of strings over a finite alphabet of size v. Define a function $k : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ as the output of running algorithm 1 on a pair of strings x, y. Show that $k(x, y)$ is a valid kernel.

3 Gaussian kernel

Problem 7: Can any *finite* set of points be linearly separated in the feature space of the Gaussian kernel

$$k_{\mathrm{G}}(x_1, x_2) = \exp\left(-\frac{|x_1 - x_2|^2}{2\sigma^2}\right),$$

if σ can be chosen freely?