Machine Learning, Tutorial 5
University of Bern

Abdelhak Lemkhenter, abdelhak.lemkhenter@inf.unibe.ch

23/10/2019

SVM

1. Suppose $K_1(x, y)$ and $K_2(x, y)$ are valid kernels. Show that each of the following items are valid kernels.

   (a) $K(x, y) = \alpha K_1(x, y)$ where $\alpha \geq 0$

   (b) $K(x, y) = K_1(x, y) + K_2(x, y)$.

   (c) $K(x, y) = \alpha K_1(x, y) + \beta K_2(x, y)$ where $\alpha, \beta \geq 0$

   (d) $K(x, y) = K_1(x, y)K_2(x, y)$

   (e) $K(x, y) = g(x)g(y)$ for $g : \mathbb{R}^n \to \mathbb{R}$

   (f) $K(x, y) = (x^\top y + d)^d$ (polynomial kernel)

   (g) $\exp\left(\frac{-||x-y||^2}{\sigma^2}\right)$ (radial basis function kernel)

   Solution. $K_1(x, y) = \phi_1(x)^\top \phi_1(y), K_2(x, y) = \phi_2(x)^\top \phi_2(y)$.

   (a)

$$\alpha K_1(x, y) = \alpha \phi_1(x)^\top \phi_1(y) = (\sqrt{\alpha}\phi_1(x)^\top)(\sqrt{\alpha}\phi_1(y))$$

   (b) $\begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix}$ is the column vector obtained by concatenating $\phi_1(x)$ and $\phi_2(x)$.

$$K(x, y) = \phi_1(x)^\top \phi_1(y) + \phi_2(x)^\top \phi_2(y) = \begin{bmatrix}\phi_1(x)\\ \phi_2(x)\end{bmatrix}^\top \begin{bmatrix}\phi_1(y)\\ \phi_2(y)\end{bmatrix} = \psi(x)^\top \psi(y)$$

   (c) Follows from (a) and (b)

   (d)

$$\begin{aligned} K_1(x, y)K_2(x, y) = &\sum_{i=1}^{n} \phi_{1i}(x)\phi_{1i}(y) \sum_{j=1}^{m} \phi_{2j}(x)\phi_{2j}(y) = \\ &\sum_{i=1}^{n}\sum_{j=1}^{m}(\phi_{1i}(x)\phi_{2j}(x))(\phi_{1i}(y)\phi_{2j}(y)) = \\ &\sum_{k=1}^{mn} \phi_{12k}(x)\phi_{12k}(y) = \phi_{12}(x)^\top \phi_{12}(y) \end{aligned}$$

   (e) set $\phi(x) = g(x)$

   (f)

$$K(x, y) = \sum_{s=0}^{d} \begin{bmatrix} d \\ s \end{bmatrix} \alpha^{d-s}(x^\top y)^s$$

   $x^\top y$ is a kernel. using (d), (b) and (a) the proof simply follows.

(g)

$$
\begin{aligned}
K(x,y) =& \exp\left(\frac{-||x-y||^2}{\sigma^2}\right) = \exp\left(\frac{-||x||^2||-y||^2+2x^\top y}{\sigma^2}\right)\\
=& \exp\left(\frac{-||x||^2}{\sigma^2}\right)\exp\left(\frac{-||y||^2}{\sigma^2}\right)\exp\left(\frac{2x^\top y}{\sigma^2}\right)\\
=& g(x)g(y)\sum_{i=0}^{\infty}\frac{2}{\sigma^2 i!}(x^\top y)^i
\end{aligned}
$$

We can see that RBF kernel is formed by taking an infinite sum over polynomial kernels.

2. Consider the constrained minimization problem below. Solve it using the KKT conditions.

$$
\begin{aligned}
\min_{x,y} \quad & x^2+y^2\\
\text{subject to} \quad & (x-3)^2+y^2 \leq 4
\end{aligned}
$$

Solution

The Lagrangian of the problem is $L(\alpha, x, y) = x^2 + y^2 + \alpha((x-3)^2 + y^2 - 4)$. The KKT conditions are:

$$
\begin{aligned}
\text{primal feasibility:} \quad & (x-3)^2+y^2 \leq 4\\
\text{dual feasibility:} \quad & \alpha \geq 0\\
\text{complementary slackness:} \quad & \alpha((x-3)^2+y^2-4)=0\\
\text{gradient of Lagrangian vanishes:} \quad & \tfrac{\partial L}{\partial x} = 2x + 2\alpha(x-3) = 0\\
& \tfrac{\partial L}{\partial y} = 2y + 2\alpha y = 0
\end{aligned}
$$

Because $1+\alpha \geq 1+0 > 0$, the only way to satisfy the last equation $(1+\alpha)y = 0$, if we set $y = 0$. According to the complementary slackness, either α or the other term is 0. If $\alpha = 0$, then $x = 0$ (from the 4^{th} equation). When $x = 0$ and $y = 0$, the primal feasibility is not satisfied, therefore $\alpha > 0$, therefore $(x-3)^2 - 4 = 0$. This leads to $x = 1$ or $x = 5$, but only $x = 1$ is feasible. The solution is therefore $x = 1$ and $y = 0$.

Interpretation of the complementary slackness condition for convex optimization The general problem we are trying to solve is the given by equation 1 where f and h are convex functions. Let's note v^* and $\hat{v}^*$ the solutions of the unconstrained and constrained problem respectively. We define the feasible set as $\{v|h(v) \leq 0\}$.

$$
\begin{aligned}
&min_v f(v)\\
&\text{s.t. } h(v) \leq 0
\end{aligned} \tag{1}
$$

We have two case. 1) $h(v^*) \leq 0$. In this case, v^* is in the feasible set so $\hat{v}^*$ should be equal to v^*. Therefore h should have no effect on the optimum of the Lagrangian, i.e α = 0. 2) $h(v^*) > 0$. Therefore v^* is not a feasible solution, since the feasible set is convex, therefore $\hat{v}^*$ will be located on the boundary of the feasible set, i.e. $h(\hat{v}^*)$. In both cases, the complementary slackness condition is satisfied.

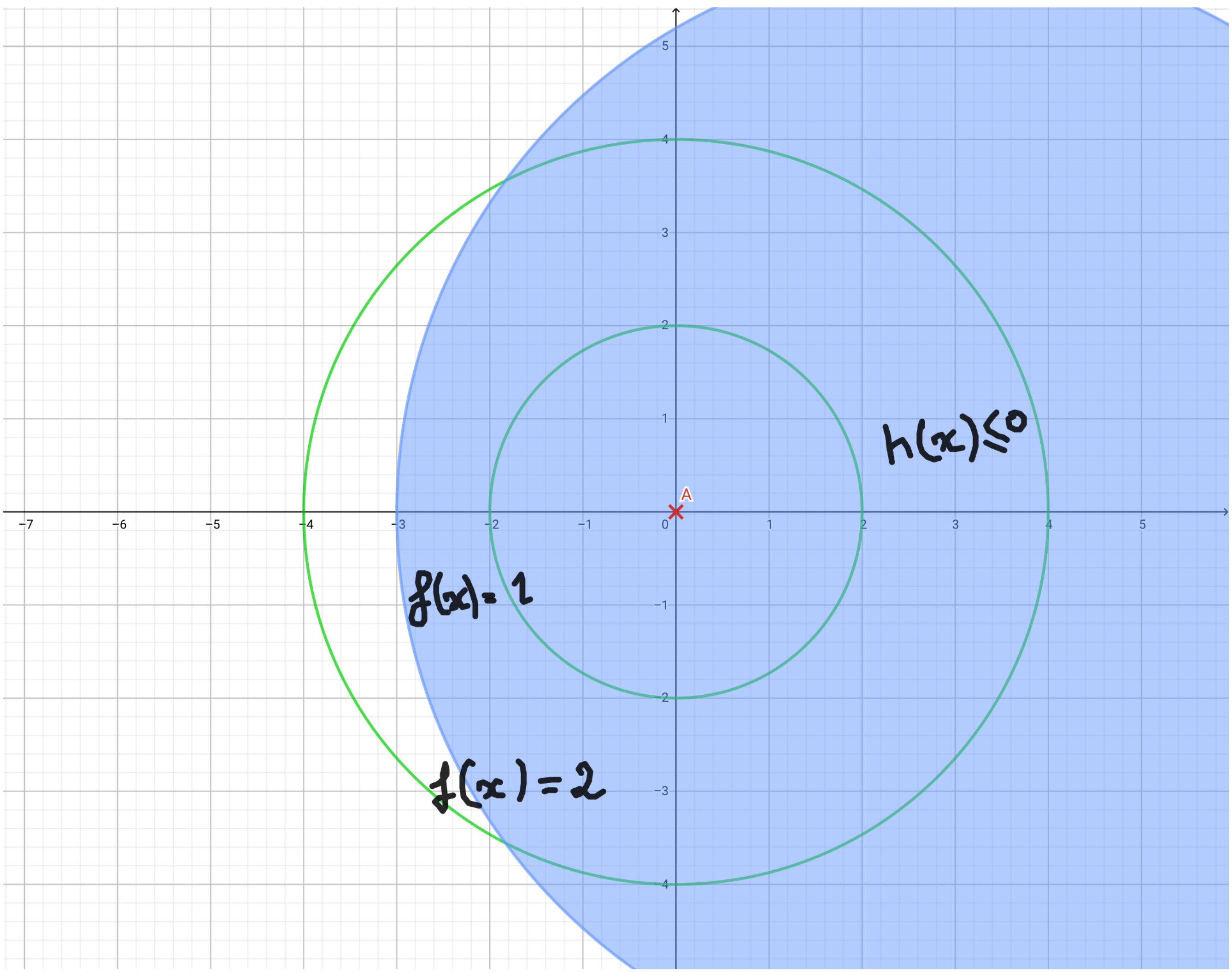

Figure 1: Illustration of case 1: $h(x,y) = ((x-3)^2 + y^2 - 36)$. $v^* = \hat{v}^* =$ point A

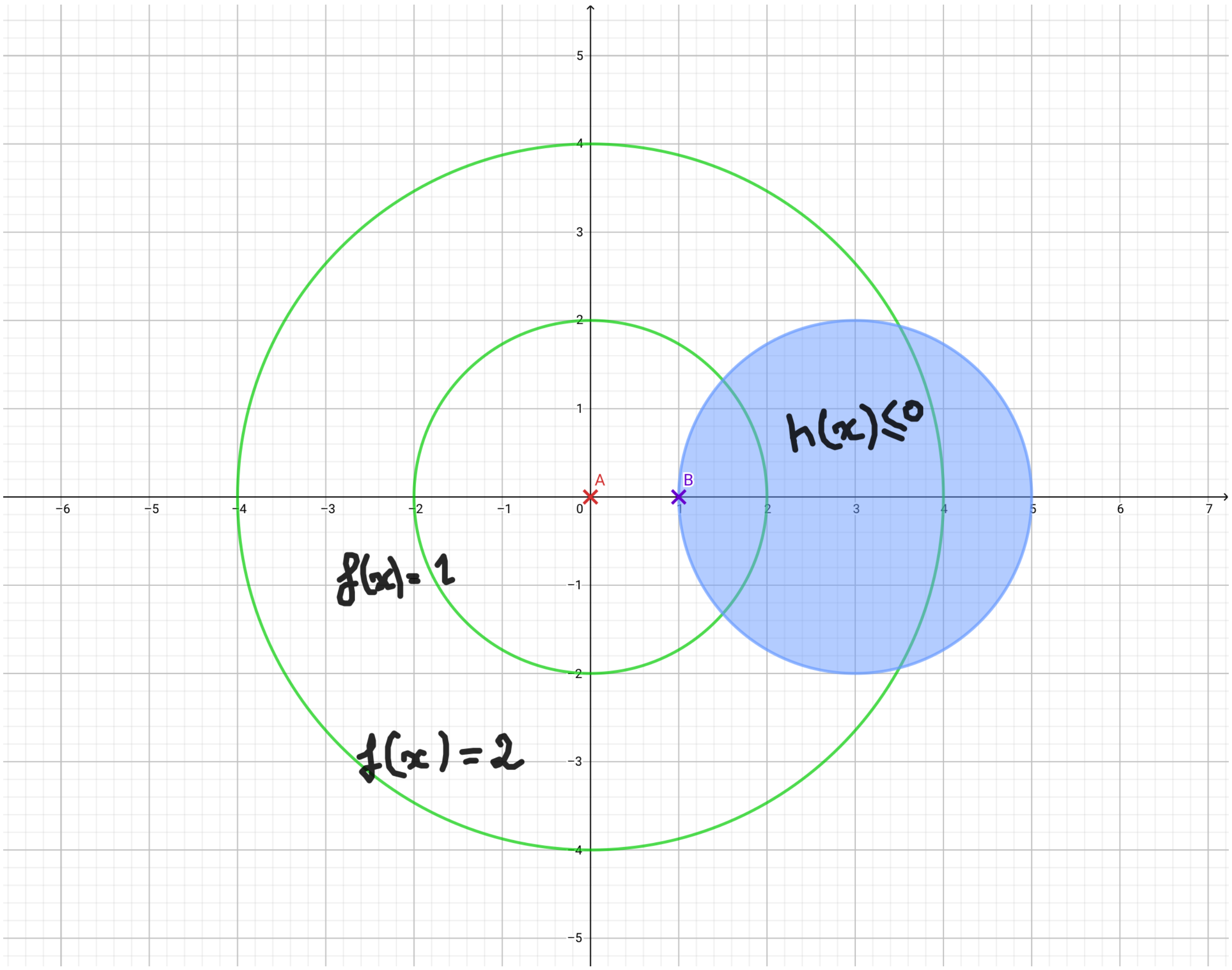

Figure 2: Illustration of case 2: $h(x, y) = ((x-3)^2 + y^2 - 4)$. v^*= point A, $\hat{v}^* =$ point B

3. Consider the following data set

- Suppose we use the SVM with a polynomial kernel for classification. What is the minimum degree of polynomial to achieve 0 training error? Explain your reasoning. **solution.** Four .Because mapping function has to cut x axis at least four times in the boundary of classes.
- Determine the four most probable support vectors with your suggested kernel.
  solution.