Data Analysis PI
Theoretical assignment #6

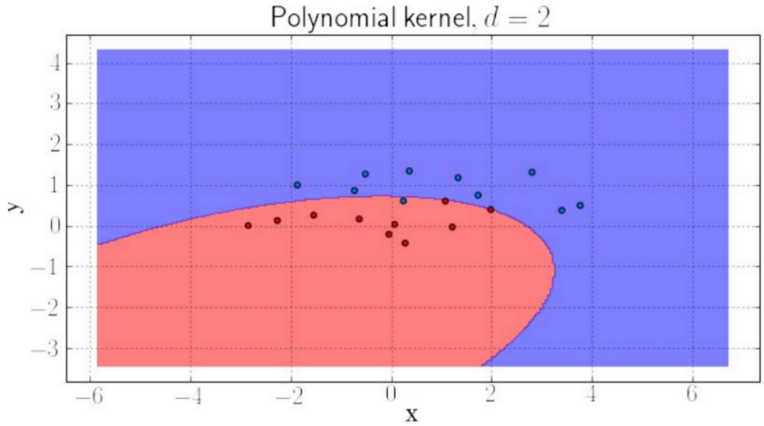
Kupriyanov Kirill

Task 1.

<u>*Problem:*</u> Consider a linearly separable dataset and SVM with polynomial kernel $K(x, y) = (x^\top y + 1)^d$. Is that right, that for any $d > 1$ the decision boundary representation in initial feature space will be the **error - free hyperplane** (linear decision boundary).

<u>*Solution:*</u> The decision boundary just cannot be linear if the kernel is RBF and $d > 1$.

For example,



Task 2.

<u>*Problem:*</u> Find the computational complexity of the linear and kernel SVM **classification** procedure of a single object (SVM is already trained).

<u>*Solution:*</u> If an input feature vector is a vector of float values, then the result is

$$Y = f(wx) = f\left(\sum_{i=1}^{n} w_i x_i\right)$$

Where w is a weight matrix, f - function to compute the result from dot product.

So, the space complexity of classifying 1 object using a linear classifier is equal to complexity of computing dot product of 2 vectors.

$$O(f) = O(d)$$

,d - number of features.

SVC:

$$a(x) = sign(\sum_{i=1}^{n} \lambda i \cdot ci \cdot xi \cdot x - b)$$

Where:

x_i - a sample object.

x - a classidied object.

Using only those objects, for which $\lambda \neq 0$, let's assume that m is number of objects for which $\lambda \neq 0$. Then, the complexity will be $O(md)$. d - number of features.

So, the linear classifier is $O(d)$, and SVM is $O(md)$

Task 3.

<u>*Problem:*</u> Consider SVM regression optimization problem. Write down dual formulation of that problem with Lagrangian and Karush-Kuhn-Takker theorem.

<u>*Solution:*</u> ε - insensitive error function

$$E_\varepsilon(a(x)-y) = \begin{cases} 0 & \text{if } |a(x)-y| < \varepsilon \\ |a(x)-y| - \varepsilon & \text{else} \end{cases}$$

The goal is to minimize a regularized error function given by

$$C\sum_{i=1}^{l} E_\varepsilon(a(x_i)-y_i) + \frac{1}{2C}||w||^2$$

Let the slacks variables be:

$$\xi_i^+ = (\langle w, x_i\rangle - w_0 - y_i - \varepsilon)_+ \geqslant 0$$
$$\xi_i^- = (-\langle w, x_i\rangle - w_0 + y_i - \varepsilon)_+ \geqslant 0$$

SVR error function:

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{l}(\xi_i^+ + \xi_i^-) \to min_{w,w_0,\xi^+,\xi^-}$$

The Lagrange multipliers are $a_n \geqslant 0, \hat{a}_n \geqslant 0, \mu_n \geqslant 0$ and $\hat{\mu} \geqslant 0$

$$L = C\sum_{i=1}^{l}(\xi_i^+ + \xi_i^-) + \frac{1}{2}||w||^2 - \sum_{i=1}^{l}(\mu_i\xi_i^+ + \hat{\mu}\xi_i^-) - \sum_{i=1}^{l} a_i(\varepsilon + \xi_i^+ + a_i - y_i) - \sum_{i=1}^{l}\hat{a}_i(\varepsilon + \xi_i^- - a_i + y_i)$$

$$\frac{dL}{dw} = 0 \to w = \sum_{i=1}^{l}(a_i - \hat{a}_i)x_i$$
$$\frac{dL}{db} = 0 \to \sum_{i=1}^{l}(a_i - \hat{a}_i) = 0$$
$$\frac{dL}{d\xi_i^+} = 0 \to a_i + \mu_i = C$$
$$\frac{dL}{d\xi_i^-} = 0 \to \hat{a}_i + \hat{\mu}_i = C$$

So the dual problem:

$$L(a,\hat{a}) = -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(a_i - \hat{a}_i)(a_j - \hat{a}_j)\varphi(x_i, x_j) - \varepsilon\sum_{i=1}^{l}(a_i + \hat{a}_i) + \sum_{i=1}^{l}(a_i - \hat{a}_i)y_i$$