AISEC Unit 3

Lecture "Machine Learning for Computer Security"
Prof. Dr. Christian Wressnegger
Artificial Intelligence & Security
Karlsruher Institute für Technologie (KIT)

1. Let us consider a set of objects $X = \{x_1, \ldots, x_n\}$ and a kernel $k : X \times X \to \mathbb{R}$ inducing a map $\phi : X \to \mathcal{F}$ to a feature space $\mathcal{F}$. The center of mass μ of the set in $\mathcal{F}$ is given by

$$\mu = \frac{1}{n} \sum_{j=1}^{n} \phi(x_j).$$

Show how the squared Euclidean distance from some object z to the center μ can be calculated using k but without using ϕ directly.
Hint: Don't confuse the number of objects (n) with the dimension of $\mathcal{F}$!

Solution:

2. Receiver Operating Characteristic (ROC) Curves
   (a) What does a ROC curve show? What does one point on the ROC curve represent?

Solution:

(b) Draw a ROC curve that has a bounded AUC (fp=$20\,\%$) of 0.5 (after normalizing to the bound). What is the value of the unbounded AUC of the same ROC curve?

Solution:

3. Develop a spam classifier that uses a polynomial bag-of-words kernel. Given an unknown message z the classifier computes the distance to the center of spam messages μ_s and the center of non-spam messages μ_h in the training data.

   Compute ROC curves for $d = \{1, 2, 3, 4\}$ and three different detection functions:

   (a) Classic anomaly detection: $f_1(z) = ||\phi(z) - \mu_h||^2$

   (b) Reverse anomaly detection: $f_2(z) = -||\phi(z) - \mu_s||^2$

   (c) Simple classification: $f_3(z) = ||\phi(z) - \mu_h||^2 - ||\phi(z) - \mu_s||^2$

   Generate one plot for each detection function comparing the different parametrizations and one additional plot comparing the best parametrization for each detection function. *Carefully label the axis, provide legends, and interpret the results.*

   Download the training and test data from ILAS: `exercises/ex03-data.zip`

Solution: