

# Exercise Sheet 5 - Kernels

## Prof. Dr. Paul Swoboda

---

**NOTE:** For this exercise sheet, you must submit a jupyter notebook and a PDF.

### Exercise 1: LogSumExp Trick (1.5 Points)

Computations with probabilities of widely varying orders of magnitude can lead to numerical instability issues with floating-point arithmetics. Such instability issues can be mildened by performing computations in log-space. Define

$$p_* = \prod_{i=1}^n p_i$$

$$p_+ = \sum_{i=1}^n p_i$$

where  $p_1, \dots, p_n \in [0, 1]$  are probabilities.

- Suppose we want to compute  $p_*$  or  $p_+$  given the  $p_i$ 's. Why can such an approach lead to numerical instabilities with floating-point arithmetics? Try to construct concrete examples.
- An alternative is to operate in log-space. Show how to compute  $\ln p_*$  given the  $\ln p_i$ 's. (You are not allowed to compute  $\exp(\ln p_i)$ , of course.)
- To compute  $\ln p_+$  given the  $\ln p_i$ 's, we may use

$$\ln p_+ = \ln \sum_{i=1}^n \exp(\ln p_i). \quad (0.1)$$

Such an approach obviously does not help to address numerical stability issues.

- The LogSumExp trick sets  $c = \max_i \ln p_i$  and uses  $\exp(\ln p_i - c)$  instead of  $\exp(\ln p_i)$ . Rewrite Eq. (0.1) using this trick (and make sure that it stays correct).
- Why does this approach reduce numerical stability issues?

## Exercise 2: Prove that Gaussian Kernel is Mercer kernel (1.5 Point)

Prove that the Gaussian kernel, defined as

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right),$$

is a Mercer kernel.

**Hint:** Use the Taylor series expansion of the exponential function and multinomial theorem to calculate the feature map of the Gaussian kernel.

## Exercise 3: Logistic Regression with kernel machine on the Two Moons Dataset (2 Points)

Using the provided implementation of logistic regression in notebook `e05_KernelLogisticRegression.ipynb`, apply a kernel machine with 20 centroids to features on the "Two Moons" dataset. Use the RBF kernel and find parameter  $\gamma$  which produces better results on the test set in terms of accuracy than logistic regression on raw Two Moons input (without kernel machine). Apply 1000 iterations.