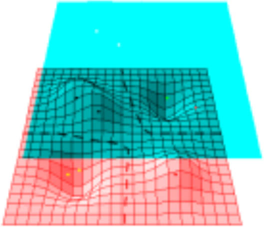
Introduction to Machine Learning

Nonlinear Support Vector Machines
The Gaussian RBF Kernel



Learning goals

- Know the Gaussian (RBF) kernel
- Understand that all data sets are separable with this kernel
- Understand the effect of the kernel hyperparameter σ

Learning goals

- Know the Gaussian (RBF) kernel
- Understand that all data sets are separable with this kernel
- Understand the effect of the kernel hyperparameter σ

RBF KERNEL / 2

- With a Gaussian kernel, all RKHS basis functions $\phi(\mathbf{x}) = k(\mathbf{x}, \cdot)$ are linearly independent - which we will not prove here.
- This means that all (finite) data sets are linearly separable!
- Do we then need soft-margin machines? The answer is "yes". The roles of the nonlinear feature map and the soft-margin constraints are very different:
  - The purpose of the kernel (and its feature map) is to make learning "easy".
  - Even in an infinite-dimensional feature space we may want some margin violators because we should not trust noisy data. A hard-margin SVM with Gaussian kernels may be able to separate any dataset but will usually overfit.

RBF KERNEL WIDTH / 2

A small σ parameter makes the function more “wiggly”, in the limit we totally over fit the data by basically modelling each training data point - and maximal uncertainty at all other test points.