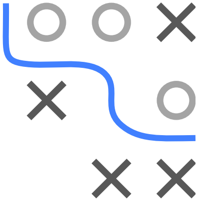
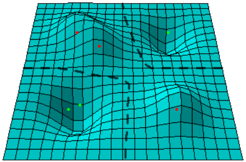
Introduction to Machine Learning



Nonlinear Support Vector Machines
The Gaussian RBF Kernel



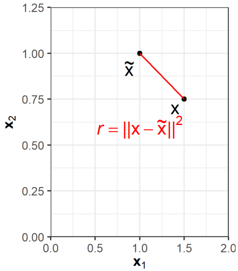
Learning goals

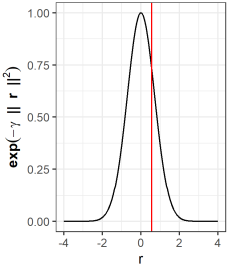
- Know the Gaussian (RBF) kernel
- Understand that all data sets are separable with this kernel
- Understand the effect of the kernel hyperparameter σ

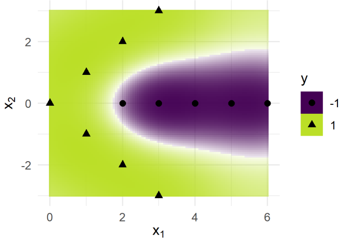
RBF KERNEL

The “radial” **Gaussian kernel** is defined as

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = \exp\left(-\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|^2}{2\sigma^2}\right) \quad \text{or} \quad k(\mathbf{x}, \tilde{\mathbf{x}}) = \exp(-\gamma\|\mathbf{x} - \tilde{\mathbf{x}}\|^2)$$





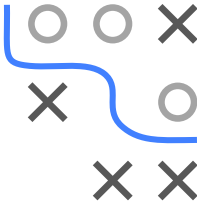


A straightforward extension is

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = \exp\left(-(\mathbf{x} - \tilde{\mathbf{x}})^T C(\mathbf{x} - \tilde{\mathbf{x}})\right)$$

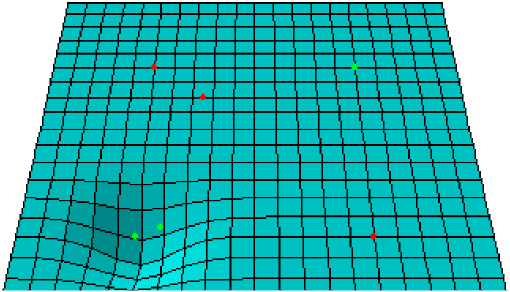
for a symmetric, positive definite matrix C.

WEIGHTED MIXTURE OF GAUSSIANS



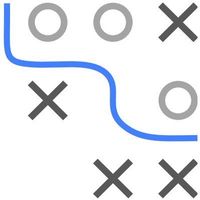
Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



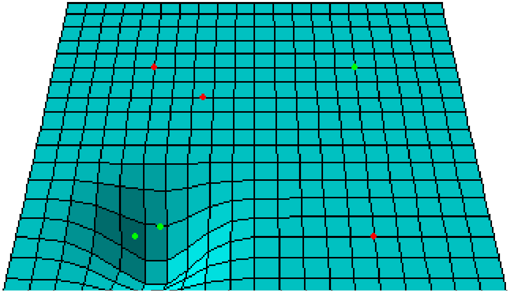
All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

WEIGHTED MIXTURE OF GAUSSIANS



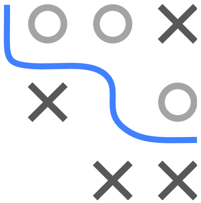
Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



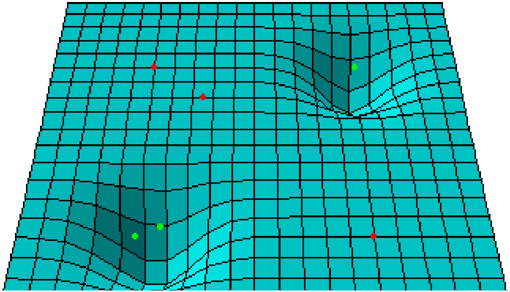
All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

WEIGHTED MIXTURE OF GAUSSIANS



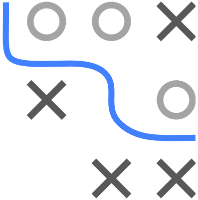
Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



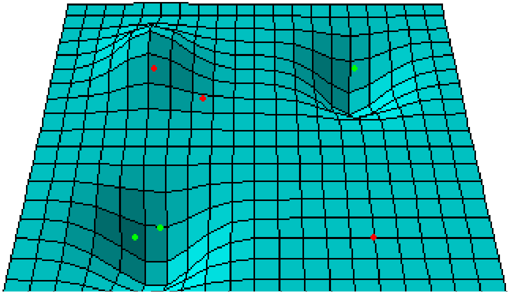
All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

WEIGHTED MIXTURE OF GAUSSIANS



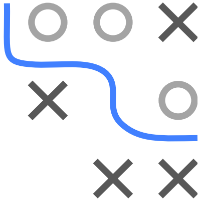
Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



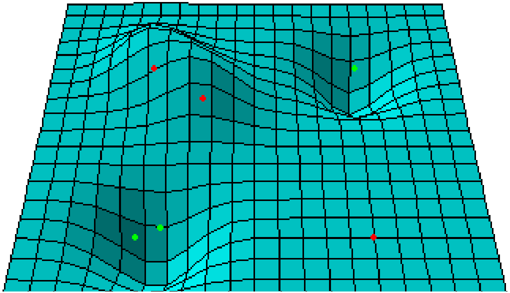
All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

WEIGHTED MIXTURE OF GAUSSIANS



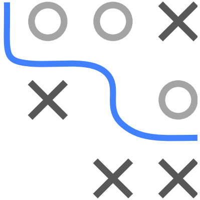
Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



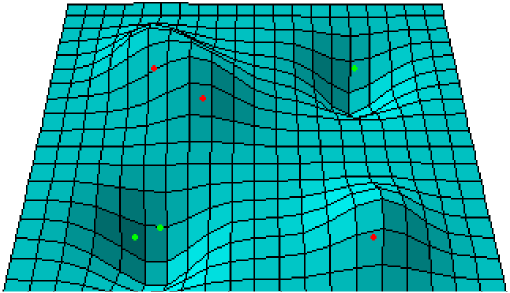
All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

WEIGHTED MIXTURE OF GAUSSIANS



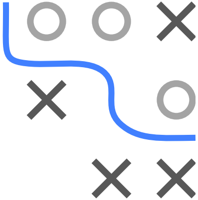
Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



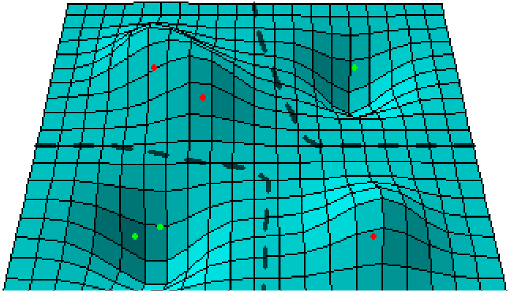
All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

WEIGHTED MIXTURE OF GAUSSIANS



Via the RKHS / basis function intuition we can understand the effect of the RBF kernel much better as a local model.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x}) + \theta_0$$



All support vectors are assigned RBF "bumps", these are weighted with the dual variables / Lagrange multipliers α_i and labels $y^{(i)}$. We then "mix" these bumps together to form the decision score function. Which becomes a bumpy surface.

RBF KERNEL WIDTH

A large σ (or a small γ) will make the decision boundary very smooth and in the limit almost linear.