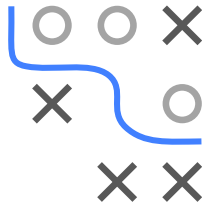
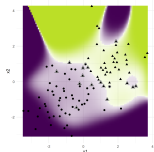


# Introduction to Machine Learning

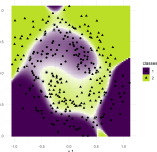
## Nonlinear Support Vector Machines The Polynomial Kernel



svm: kernelpolynomial, degree=2, coef0=1  
Train: mince=0.170000, CV: mince: test: mean=0.250000



svm: kernelpolynomial, degree=3, coef0=1  
Train: mince=0.059999, CV: mince: test: mean=0.120000



### Learning goals

- Know the homogeneous and non-homogeneous polynomial kernel
- Understand the influence of the choice of the degree on the decision boundary

# HOMOGENEOUS POLYNOMIAL KERNEL

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = (\mathbf{x}^T \tilde{\mathbf{x}})^d, \text{ for } d \in \mathbb{N}$$

The feature map contains all monomials of exactly order  $d$ .

$$\phi(\mathbf{x}) = \left( \sqrt{\binom{d}{k_1, \dots, k_p}} x_1^{k_1} \dots x_p^{k_p} \right)_{k_i \geq 0, \sum_i k_i = d}$$

That  $\langle \phi(\mathbf{x}), \phi(\tilde{\mathbf{x}}) \rangle = k(\mathbf{x}, \tilde{\mathbf{x}})$  holds can easily be checked by simple calculation and using the multinomial formula

$$(x_1 + \dots + x_p)^d = \sum_{k_i \geq 0, \sum_i k_i = d} \binom{d}{k_1, \dots, k_p} x_1^{k_1} \dots x_p^{k_p}$$

The map  $\phi(\mathbf{x})$  has  $\binom{p+d-1}{d}$  dimensions. We see that  $\phi(\mathbf{x})$  contains no terms of "lesser" order, so, e.g., linear effects. As an example for  $p = d = 2$ :  $\phi(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ .



# NONHOMOGENEOUS POLYNOMIAL KERNEL

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = (\mathbf{x}^T \tilde{\mathbf{x}} + b)^d, \text{ for } b \geq 0, d \in \mathbb{N}$$

The maths is very similar as before, we kind of add a further constant term in the original space, with

$$(\mathbf{x}^T \tilde{\mathbf{x}} + b)^d = (x_1 \tilde{x}_1 + \dots + x_p \tilde{x}_p + b)^d$$

The feature map contains all monomials up to order  $d$ .

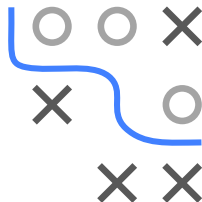
$$\phi(\mathbf{x}) = \left( \sqrt{\binom{d}{k_1, \dots, k_{p+1}}} x_1^{k_1} \dots x_p^{k_p} b^{k_{p+1}/2} \right)_{k_i \geq 0, \sum_i k_i = d}$$

The map  $\phi(\mathbf{x})$  has  $\binom{p+d}{d}$  dimensions. For  $p = d = 2$ :

$$(x_1 \tilde{x}_1 + x_2 \tilde{x}_2 + b)^2 = x_1^2 \tilde{x}_1^2 + x_2^2 \tilde{x}_2^2 + 2x_1 x_2 \tilde{x}_1 \tilde{x}_2 + 2bx_1 \tilde{x}_1 + 2bx_2 \tilde{x}_2 + b^2$$

Therefore,

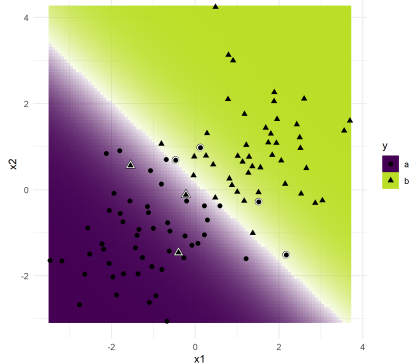
$$\phi(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1 x_2, \sqrt{2b}x_1, \sqrt{2b}x_2, b)$$



# POLYNOMIAL KERNEL

Degree  $d = 1$  yields a linear decision boundary.

svm: kernel=polynomial; degree=1; coef0=1  
Train: mmce=0.0700000; CV: mmce.test.mean=0.1100000



svm: kernel=polynomial; degree=1; coef0=1  
Train: mmce=0.5000000; CV: mmce.test.mean=0.5066667

