Feature Maps and Kernels
Last Time

- SVM primal problem has a dual optimization
- Dual has box constraints on dual variables
- Dual only considers inner products of data vectors
- Kernel trick: replace inner products with kernel functions
  - Inner products in mapped feature space
Kernel SVM

$$\min_{\alpha} \frac{1}{2} \sum_{i} \sum_{j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i} \alpha_i$$

s.t. \( \sum_{i} \alpha_i y_i = 0, \quad \alpha_i \in [0, C] \)

$$f(x) = w^\top x + b = \sum_{i} \alpha_i y_i K(x_i, x) + b \quad b = y_i - \sum_{j} \alpha_j y_j K(x_i, x_j)$$

for examples \( i \) where \( 0 < \alpha_i < C \)

\( K = \) kernel function
Outline

• Feature maps and nonlinearity
• Efficient kernel functions
  • Polynomial kernel
• Gaussian radial basis function kernel (correction from last video)
Nonlinear Decision Boundary

$wx - b$
Nonlinear Decision Boundary

\[ wx - b \]
Nonlinear Decision Boundary
Nonlinear Decision Boundary
Polynomial Feature Map

\[ \Phi(x) = [x^1, \ldots, x^d, x^1x^1, \ldots, x^1x^d, \ldots, x^dx^1, \ldots, x^dx^d, \ldots]^{\top} \]

Third-order terms \( \{x^1x^1x^1, x^1x^1x^2, \ldots, x^1x^dx^d, \ldots, x^dx^dx^d\} \)

Fourth-order terms \( \{x^ix^jx^kx^\ell | i, j, k, \ell \in \{1, \ldots, d\}\} \)

\[ |\Phi(x)| = \sum_{a=1}^{M} d^{a} = O(d^{M}) \]
Polynomial Decision Scores

- Quadratic
- Quadratic
- Third order
Efficient Kernel Computation

\[ K(x_i, x_j) = \Phi(x_i)^\top \Phi(x_j) \]

\[ \Phi(x_i) = [x_i^1, \ldots, x_i^d, x_i^1 x_i^1, \ldots, x_i^1 x_i^d, \ldots, x_i^d x_i^1, \ldots, x_i^d x_i^d, \ldots]^\top \]

\[ \Phi(x_j) = [x_j^1, \ldots, x_j^d, x_j^1 x_j^1, \ldots, x_j^1 x_j^d, \ldots, x_j^d x_j^1, \ldots, x_j^d x_j^d, \ldots]^\top \]

\[ (x_i^\top x_j + 1)^M \]

\[ (x_i^\top x_j)(x_i^\top x_j) + 2x_i^\top x_j + 1 \]

\[ K = (X^\top X + 1)^M \]

elementwise exponentiation
Radial Basis Functions
RBF Kernel

\[ K(x_i, x_j) = \exp \left( -\frac{1}{2\sigma^2} \|x_i - x_j\|^2 \right) \]

(mistake from last video)

\[ \Phi(x) = \left[ \exp \left( \frac{1}{\sigma} \|x - x_1\|^2 \right), \exp \left( \frac{1}{\sigma} \|x - x_2\|^2 \right), \ldots, \exp \left( \frac{1}{\sigma} \|x - x_n\|^2 \right) \right] \]

What is \( \Phi(x) \)?
Taylor Expansion of RBF Kernel

\[ K(x_i, x_j) = \exp \left( -\frac{1}{2\sigma^2} \|x_i - x_j\|^2 \right) \]

\[ = \exp \left( -\|x_i - x_j\|^2 \right) \]

\[ = \exp (-x_i^\top x_i) \exp (-x_j^\top x_j) \exp(2x_i^\top x_j) \]

\[ = \exp (-x_i^\top x_i) \exp (-x_j^\top x_j) \sum_{n=0}^{\infty} \frac{2^n(x_i^\top x_j)^n}{n!} \]

\[ \Phi^{r\text{bf}} = \exp(-x^\top x) \left[ \Phi^1(x)^\top, \Phi^2(x)^\top, \ldots, \Phi^\infty(x)^\top \right]^\top \]
Kernel Formulas

**Linear**

\[ K(x_i, x_j) = x_i^\top x_j \quad \text{with} \quad X_i \in \mathbb{R}^{d \times m}, \quad X_j \in \mathbb{R}^{d \times n} \]

\[ K = X_i^\top X_j \]

**Polynomial**

\[ K(x_i, x_j) = (x_i^\top x_j + 1)^M \]

\[ K = (X_i^\top X_j + 1)^M \]

**RBF**

\[ K(x_i, x_j) = \exp \left( -\frac{1}{2\sigma^2} \|x_i - x_j\|^2 \right) \]

\[ K = \exp \left( -\frac{1}{2\sigma^2} \left( \text{diag}(X_i^\top X_i)^\top + \text{diag}(X_j^\top X_j)^\top - 2X_i^\top X_j \right) \right) \]
Kernels

• Map input data to new feature space (usually higher dimensional)
• Efficient method for computing inner product in mapped space
• Methods using inner products can directly use kernel
  • E.g., dual SVM
Next Time

• How to alleviate the computational cost of SVM training?

• QP: roughly $O(n^3)$ for $n$ constraints or variables