Discussion 6: Kernels
Machine Learning, Spring 2019

Properties

(a) (1) If k_1 and k_2 are valid kernels,
$k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z}) k_2(\mathbf{x}, \mathbf{z})$

Solution Based on Mercer's Theorem, we have $k_1(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{\infty} \lambda_i \psi_i(\mathbf{x}) \psi_i(\mathbf{z})$, similar form for $k_2(\mathbf{x}, \mathbf{z})$ can be obtained. Therefore,

$$\begin{aligned} k(\mathbf{x}, \mathbf{z}) &= \sum_{i=1}^{\infty} \lambda_i \psi_i(\mathbf{x}) \psi_i(\mathbf{z}) \sum_{j=1}^{\infty} \lambda_j \phi_j(\mathbf{x}) \phi_j(\mathbf{z}) \\ &= \sum_{i,j} \lambda_i \lambda_j [\psi_i(\mathbf{x}) \phi_j(\mathbf{x})][\psi_i(\mathbf{z}) \phi_j(\mathbf{z})] \\ &= \langle \Upsilon(\mathbf{x}), \Upsilon(\mathbf{z}) \rangle_{\mathcal{H}_{new}} \end{aligned}$$

where $\Upsilon(\mathbf{x}) = [..., \sqrt{\lambda_i \lambda_j} \psi_i(\mathbf{x}) \phi_j(\mathbf{x}), ...]$
(2) $k(\mathbf{x}, \mathbf{z}) = g(\mathbf{x}) g(\mathbf{z})$ for $g : \mathcal{X} \to \mathbb{R}$.

Solution Based on the definition of kernel, we need to prove the following two properties:
1) Symmetric: Obviously $k(\mathbf{x}, \mathbf{z}) = k(\mathbf{z}, \mathbf{x})$
2) Given any dataset $\mathbf{x}_i$, $i = 1, ..., N$, define vector $\mathbf{g} = (g(\mathbf{x}_1), g(\mathbf{x}_2), ..., g(\mathbf{x}_N))^\top$.

$$\mathbf{K} = \mathbf{g}\mathbf{g}^\top$$

then $\mathbf{K}$ is PSD, since $\forall \mathbf{v}, \mathbf{v}^\top \mathbf{K} \mathbf{v} = (\mathbf{v}^\top \mathbf{g})^2 \geq 0$

(b) Assume there exists some matrix B with negative eigenvalues. Show that B cannot be used to define an inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathbf{T}} B \mathbf{y}$.

Solution Let λ be a negative eigenvalue with corresponding eigenvector $\mathbf{v}$: $K\mathbf{v} = \lambda \mathbf{v}$, then:

$$\mathbf{v}^T K \mathbf{v} = \lambda \mathbf{v}^T \mathbf{v} = \lambda \mathbf{v}^T \mathbf{v} < 0$$

Since the inner product $\mathbf{v}^T K \mathbf{v}$ is required to be greater than zero, a contradiction has resulted.

(c) Let $\mathbf{A}$ be a positive semi-definite (PSD) matrix. Show that an inner product defined as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{A} \mathbf{y}$ may not be valid.

Solution We can give a counterexample,

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

$\forall \mathbf{x}, \mathbf{x}^T A \mathbf{x} = (x_1 + x_2)^2 \geq 0$, so it is PSD.
However, $\mathbf{x}^T A \mathbf{y} = x_1 y_1 + x_2 y_2 + 2 x_1 y_2$ which is not symmetric.

Visualizing kernels

3.1:

(a) polynomial kernel

(b) RBF kernel

(c) linear kernel

(d) sigmoid kernel

3.2:

1. upper left: CUBIC and RBF
2. upper right: RBF
3. lower left: QUADRATIC, CUBIC and RBF
4. lower right: QUADRATIC, CUBIC and RBF

SVM with a non-linear kernel

Answer: $\phi(\vec{x}) = \sqrt{2}\|\vec{x}\|$

Step 2: Convert all our original points into the new space using the transform. (We are going from 2D to 1D).
Positive points are at:

$\phi(p_1) = \sqrt{2} \cdot 0$

$\phi(p_3) = \sqrt{2} \cdot 1\sqrt{2} = 2$

$\phi(p_4) = \sqrt{2} \cdot 2\sqrt{2} = 4$

$\phi(p_6) = \sqrt{2} \cdot 1\sqrt{2} = 2$

Negative points are at:

$\phi(p_5) = \sqrt{2} \cdot 3\sqrt{2} = 6$

$\phi(p_2) = \phi(p_7) = \sqrt{2} \cdot 4\sqrt{2} = 8$

Step 3: Plot the points in the new space, this appears as a line from 0 to 8.
With positive points at 0, 2, 4 and negative points at 6, 8.

The support vectors lie between $\phi(p_4)$ and $\phi(p_5)$ (between values of 4 and 6)
Hence the decision boundary (maximum margin) should be: $\phi(x) < 5$
The < due to the positive points being all less than 5.

Expanding the determined decision boundary in terms of components of x, we get:

$$\phi(x) = \sqrt{2} \cdot \sqrt{x_1^2 + x_2^2} < 5$$

Square both sides:

$$2(x_1^2 + x_2^2) < 25$$

Convert to $\geq$ (standard form):

$$-2x_1^2 - 2x_2^2 + 25 \geq 0$$

$$(x_1^2 + x_2^2) < \frac{25}{2}$$

This is a circle with radius $5/\sqrt{2} = \frac{5}{2}\sqrt{2} = 2.5\,\text{diagonals} \approx 3.5$

Figure 1