3. Kernels

In this question, we will practice kernel methods in classification.

3. (a)

1.0/1 point (graded)

Let $x, q \in \mathbb{R}^2$ be two feature vectors, and let $K(x, q) = (x^T q + 1)^2$. This is often known as a polynomial kernel. It's simple to compute: you just take the dot product between two feature vectors, add one, and then square the result. But what kind of feature mapping does this kernel implicitly use?

Assuming we can write $K(x, q) = \phi(x)^T \phi(q)$, derive an expression for $\phi(x)$.

Enter the solution as a vector $\phi(x) = [f_1(x_1, x_2), \cdots, f_N(x_1, x_2)]$

$\phi(x) =$ [x_1^2, x_2^2, sqrt(2)*x_1*x_2, sqrt(2)*x_1,sqrt(2)*x_2,1]

Answer: [x_1^2, x_2^2, sqrt(2)*x_1*x_2, sqrt(2)*x_1, sqrt(2)*x_2, 1]

Solution:

- We can rewrite the kernel as $K(x, q) = (x^T q + 1)^2 = \left(1 + \sum_{i=1}^{2} x_i q_i\right)^2 = (x_1 q_1 + x_2 q_2 + 1)^2$.
- Expanding and combining terms gives $x_1^2 q_1^2 + x_2^2 q_2^2 + 2x_1 x_2 q_1 q_2 + 2x_1 q_1 + 2x_2 q_2 + 1$.
- We can then rewrite this expression as $\phi(x)^T \phi(q)$ where $\phi(x) = [x_1^2, x_2^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1]$

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

3. (b)

1/1 point (graded)

As a simple example that uses this kernel, imagine that our feature vectors were bag of words vectors. In this example, give an intuitive interpretation of what the $\sqrt{2}x_1 x_2$ term in the expression for $\phi(x)$ you just wrote down means.

- consecutive co-appearance (bigram)
- co-appeareance in document ✔

Solution:

- Each token in the bag-of-word model only represents appearance in the document.
- Hence, $x_1 x_2$ represents co-appearance in a document.

Submit You have used 1 of 1 attempt