Gaussian Feature Map

Chieh Wu

Northeastern University

February 2020

A Gaussian kernel is the function where given 2 samples $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^d$ we get

$$k(x, y) = e^{-\gamma||x-y||^2}. \tag{1}$$

We now write out the derivation to get its feature map.

$$\begin{aligned} k(x, y) &= e^{-\gamma||x-y||^2} & (2)\\ &= e^{-\gamma(x^T - y^T)(x-y)} & (3)\\ &= e^{-\gamma(x^T x + y^T y - 2x^T y)} & (4)\\ &= \underbrace{\left(e^{-\gamma x^T x}\right)}_{\text{Part 1}} \underbrace{\left(e^{-\gamma y^T y}\right)}_{\text{Part 2}} \underbrace{\left(e^{2\gamma x^T y}\right)}_{\text{Part 3}}. & (5) \end{aligned}$$

Notice that for part 3, we can apply the Taylor expansion of an exponential rule where

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + + \frac{x^2}{2!} + ... \tag{6}$$

Following this rule Eq. (5) becomes

$$\left(e^{-\gamma x^T x}\right) \left(e^{-\gamma y^T y}\right) \underbrace{\left(1 + (2\gamma x^T y) + \frac{(2\gamma x^T y)^2}{2!} + ...\right)}_{\text{Part 3}}. \tag{7}$$

Here, its easier to see how the terms split into the inner product of two features if we use a simple example where $x = [x_1 x_2]^T$ and $y = [y_1 y_2]^T$. Given this, part 3 becomes

$$\begin{aligned} &= \left(1 + (2\gamma(x_1 y_1 + x_2 y_2)) + \frac{(2\gamma(x_1 y_1 + x_2 y_2))^2}{2!} + ...\right) & (8)\\ &= \left(1 + (2\gamma(x_1 y_1 + x_2 y_2)) + \frac{(4\gamma^2(x_1^2 y_1^2 + x_2^2 y_2^2 + x_1 x_2 y_1 y_2))}{2!} + ...\right). & (9) \end{aligned}$$

By looking at Eq. (7) and (9), we see that the feature map for x is

$$\left(e^{-\gamma x^T x}\right) \begin{bmatrix} 1 & \sqrt{2\gamma} x_1 & \sqrt{2\gamma} x_2 & 2\gamma \frac{x_1^2}{\sqrt{2!}} & 2\gamma \frac{x_2^2}{\sqrt{2!}} & 2\gamma \frac{x_1 x_2}{\sqrt{2!}} & ... \end{bmatrix} \tag{10}$$