- $X = (x_1, \ldots, x_n)^T$ are the observations. Each x_j is a D-dimensional row vector.
- x^* test point. x^* is itself a D-dimensional row vector
- $K(X, x^*)$ kernel: $(n \times 1)$ matrix
- $\frac{\partial K(X,x^*)}{\partial x^*}$ kernel derivative: $(n \times D)$ matrix given by

$$\frac{\partial K(X, x^*)}{\partial x^*} = \frac{(X - x^*)}{\ell^2} * K(X, X^*) \tag{1}$$

  where $(X - x^*) = (x_1 - x^*, \ldots, x_n - x^*)^T$ and $*$ is the element wise product in python.

- acquisition function

$$a(z, x^*) = -\sigma(x^*) \ \left[\phi(z) + z\Phi(z)\right]$$

  where

$$z = \frac{y_{\text{best}} - \mu(x^*)}{\sigma(x^*)}$$

- acquisition function derivative

$$\frac{da(z, x^*)}{dx^*} = \Phi(z)\frac{dz}{dx^*} + \left[\phi(z) + z\Phi(x)\right] \frac{d\sigma}{dx^*}$$

  with

$$\frac{dz}{dx^*} = -\frac{1}{\sigma}\frac{d\mu}{dx^*} - \frac{y_{\text{best}} - \mu}{\sigma^2}\frac{d\sigma}{dx^*}$$