$$\therefore \phi_{rbf}(x)^T \phi_{new}(y) = \phi(x)^T \phi(y) \cdot \exp\left(\frac{-1}{2\sigma^2} x^T x\right) \cdot \exp\left(\frac{-1}{2\sigma^2} y^T y\right).$$

$$= \exp\left(\frac{x^T y}{\sigma^2}\right) \cdot \exp\left(\frac{-1}{2\sigma^2} x^T x\right) \cdot \exp\left(\frac{-1}{2\sigma^2} y^T y\right).$$

$$= \exp\left(\frac{-1}{2\sigma^2} \left(\|x - y\|^2\right)\right).$$

$$= K(x, y)$$

Hence Proved, the Gaussian (rbf) kernel can be expressed in terms of $\phi^T(x).\phi(y)$, Hence, rbf kernel is a valid kernel.

Q-1.2)b).(i). ~~Best~~ $\sigma = 1$ (minimum errors at $\sigma = 1$).

(iii) As we ~~increase~~ decrease the σ, we basically are shrinking the neighborhood of point for performing our prediction.

Upto $\sigma = 1$, the errors decrease as we are shrinking the neighborhood to consider just a small meaningful set of neighbours.

But below $\sigma = 1$, our errors increase as ~~we begin~~ our interval shrinks down tremendously and we begin to overfit on our data.

[Plot on the next page].