3) Equivalence between GPR and SVR

a) Using an rbf kernel, for what value of the parameter b in SVR, the regression value far away from the data would be the same as the mean GPR value.

b) How would the formulation in SVR change if it is required to learn the SVR function with a fixed value of $b = 0$ (Hint: What is the effect of $b$ in the dual).

c) Show that if noise is not considered in both GPR and SVR, the modified SVR above (fixed $b$) and the GP mean regression values are equivalent.

Solution

a) The GPR mean regression function is of the form $\sum_i \alpha_i k(x_i, x_j)$ (slide 62 of lecture), i.e., there is no bias term as it is in the SVR. For rbf kernel, the kernel function vanishes far away from the data and the function value is only $b$. Hence, SVR with $b = 0$ would give the same value as GPR far from the data (which is equal to zero as well).

b) In general, $b$ is a variable which is optimized for in the primal, resulting in a linear constraint on $\alpha_i$ (slide 23 of lecture), i.e., $\frac{\partial L}{\partial b} = \sum_i (\alpha_i - \alpha_i^*) = 0$. If $b$ is a fixed value, then we do not set $\frac{\partial L}{\partial b} = 0$ and hence do not get the corresponding linear constraint in the dual. The dual optimization then becomes

$$\begin{align*}
&\text{minimize} & & \frac{1}{2} \sum_{i,j} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j) + \epsilon \sum_i (\alpha_i + \alpha_i^*) - \sum_i y_i (\alpha_i - \alpha_i^*) \\
&\text{subject to} & & \alpha_i > 0 \ ; \ \alpha_i^* > 0.
\end{align*}$$

In effect, the resulting values of the Lagrange multipliers $\alpha_i$ have larger magnitudes as compared to the general SVR case, where $b$ is set to the average function value and subsequent errors due to non-linearity are accounted for by the term $\sum_i \alpha_i k(x_i, x_j)$. The attached figure illustrates this fact.

c) In the absence of noise, we would want the SVR function to pass through all the data points exactly, i.e., $\epsilon = 0$. In this case, the optimization problem above simplifies further to

$$\begin{align*}
&\text{minimize} & & \frac{1}{2} \sum_{i,j} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j) - \sum_i y_i (\alpha_i - \alpha_i^*) \\
&\text{subject to} & & \alpha_i > 0 \ ; \ \alpha_i^* > 0.
\end{align*}$$

Note that the only terms appearing in the objective are $(\alpha_i - \alpha_i^*)$. Also, since there are only positivity constraints on the Lagrange multipliers, the quantity $(\alpha_i - \alpha_i^*)$ is unconstrained. We rewrite $\alpha_i \equiv (\alpha_i - \alpha_i^*)$ and the resulting optimization is an unconstrained one:

$$\text{minimize} \ \frac{1}{2} \alpha^T K \alpha - y^T \alpha$$

where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_M]^T$ $K \equiv [k(x_i, x_j)]_{ij}$ and $y = [y_1, y_2, \ldots, y_M]^T$ is the vectorized notation.

Minimizing this objective by taking derivative with respect to $\alpha$ and setting to zero we get

$$K \alpha = y \Rightarrow \alpha = K^{-1} y.$$ 

In GPR, with no noise, i.e., $\sigma = 0$, we get (slide 61 of lecture)

$$\alpha = (K + \sigma^2 I)^{-1} y = K^{-1} y.$$ 

Hence, with no noise in the data and $b$ forced to zero with the modification in part b), GPR and SVR are equivalent.