TP1: Bayesian Linear Regression

Fondements Théoriques de l'Apprentissage Profond (MVA)

Aziz Bacha

Bayesian linear regression: results of the predictive distribution on the synthetic dataset

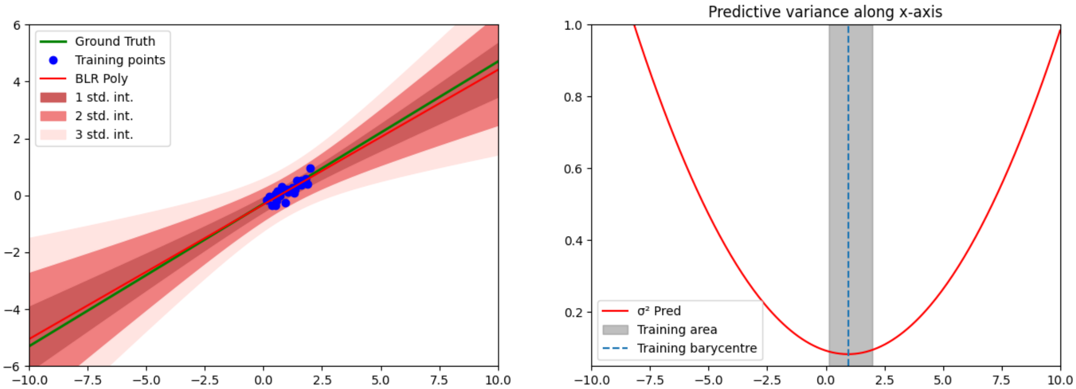

Figure 1: Bayesian Linear Regression: Predictive Mean and Uncertainty

Theoretical analysis to explain the form of the distribution

In linear regression, uncertainty in the slope parameter has little impact near the data, but this uncertainty is amplified as $|x^*|$ grows. This is why predictive uncertainty increases when extrapolating far from the training region.

For $\alpha = 0$ and $\beta = 1$, and a 1D linear basis $\phi(x) = [1,\ x]^\top$, the predictive variance is

$$\sigma^2_{\text{pred}}(x^*) = 1 + \phi(x^*)^\top (\Phi^\top \Phi)^{-1} \phi(x^*).$$

Using

$$\Phi^\top \Phi = \begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix},$$

we obtain the closed-form expression

$$\sigma^2_{\text{pred}}(x^*) = 1 + \frac{1}{N} + \frac{(x^* - \bar{x})^2}{N\,\text{Var}(x)},$$

where $\bar{x}$ and $\text{Var}(x)$ are the empirical mean and variance of the training inputs.

Finally,

$$\lim_{|x^*| \to \infty} \sigma^2_{\text{pred}}(x^*) = +\infty.$$

Thus, predictive variance grows outside the training distribution because parameter uncertainty is multiplied by the distance $|x^* - \bar{x}|$.

Non-linear regression: analysis of the Gaussian basis feature maps results

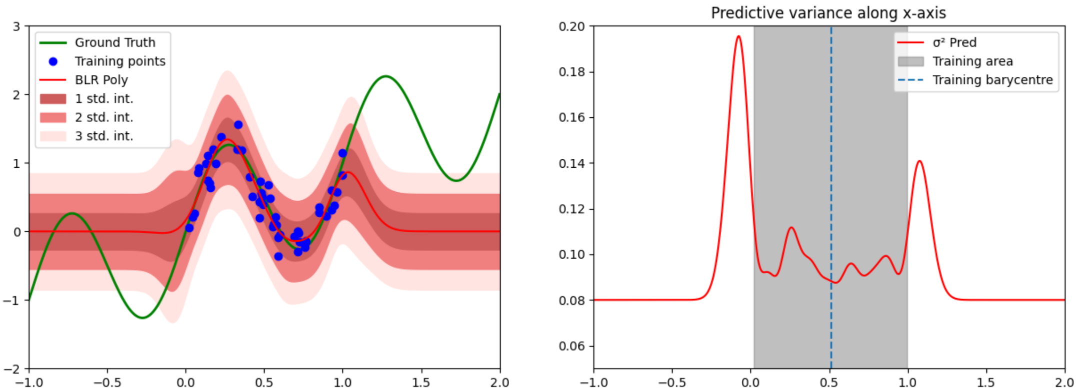

Figure 2: Bayesian linear regression with Gaussian basis functions

With the basis Gaussian function $\phi_j(x) = \exp\left(-\frac{(x - c_j)^2}{2\ell^2}\right)$, each feature goes to 0 when x is far from its center c_j. So, far from all training points: $\phi(x) \to 0$.

The predictive variance is $\sigma^2_{\text{pred}}(x) = \frac{1}{\beta} + \phi(x)^\top \Sigma\, \phi(x)$.

As $\phi(x) \to 0$, the data term $\phi(x)^\top \Sigma\, \phi(x)$ vanishes, leaving only the noise term.

Thus, outside the training distribution: $\sigma^2_{\text{pred}}(x) \longrightarrow \frac{1}{\beta}$.