Deep learning applications

Deep Learning Practical Work

Aymeric Delefosse & Charles Vin

2023 – 2024

Contents

Chapter 1

Bayesian Linear Regression

1.1 Linear Basis function model

1.1.1 Gaussian basis functions

Recall closed form of the posterior distribution in linear case. Then, code and visualize posterior sampling. What can you observe?

$$p(\mathbf{w}|\mathbf{X}, \mathbf{Y}) = \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

$$\boldsymbol{\Sigma}^{-1} = \alpha\mathbf{I} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}.$$

$$\boldsymbol{\mu} = \beta\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\mathbf{Y}.$$

Recall and code closed form of the predictive distribution in linear case.

$$p\left(y|x^*; \mathbf{D}, \alpha, \beta\right) = \mathcal{N}\left(y; \mu^T\boldsymbol{\Phi}(x^*), \frac{1}{\beta} + \boldsymbol{\Phi}(x^*)^T\boldsymbol{\Sigma}\boldsymbol{\Phi}(x^*)\right).$$

Based on previously defined `f_pred()`, predict on the test dataset. Then visualize results using `plot_results()` defined at the beginning of the notebook.

Analyse these results. Why predictive variance increases far from training distribution? Prove it analytically in the case where $\alpha = 0$ and $\beta = 1$. With $\alpha = 0, \beta = 1$, the computation of Σ^{-1} become

$$\begin{aligned}\Sigma^{-1} &= 0 * Id_3 + 1 * \phi^T\phi = \phi^T\phi \\ &= \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots 1 \quad x_N \end{pmatrix} \\ &= \begin{pmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}\end{aligned}$$

Finally we inverse Φ using the classic formula for 2×2 matrix

$$\Sigma = \frac{1}{\det \Sigma^{-1}} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & N \end{pmatrix}.$$

Bonus: What happens when applying Bayesian Linear Regression on the following dataset?

1.2 Non Linear models

1.2.1 Polynomial basis functions

Code and visualize results on sinusoidal dataset using polynomial basis functions. What can you say about the predictive variance?

1.2.2 Gaussian basis functions

Code and visualize results on sinusoidal dataset using Gaussian basis functions. What can you say this time about the predictive variance?

Explain why in regions far from training distribution, the predictive variance converges to this value when using localized basis functions such as Gaussians.