Bayesian Linear Regression — Practice Questions

CM52054: Foundational Machine Learning
Practice set with fully worked answers

Conceptual Questions

1) **Probabilistic model for linear regression.**
What is the probabilistic model assumed for linear regression?

2) **Meaning of i.i.d. noise.**
What does "i.i.d. noise" mean in this context?

3) **Likelihood for the whole dataset.**
Write down the likelihood $p(y \mid X, w)$ for the whole dataset under the Gaussian noise model.

4) **Optimisation problem from ML estimation.**
What optimisation problem does Maximum Likelihood (ML) estimation lead to in this setting?

5) **Difference between ML and MAP.**
What is the difference between ML and MAP estimation?

6) **Effect of a Gaussian prior.**
If we set a Gaussian prior $p(w) = \mathcal{N}(0, I)$, how does this affect the solution compared to ML?

Short Derivations and Calculations

7) **Normal equations for the ML estimator.**
Starting from

$$w^* = \arg\min_w \|Xw - y\|^2,$$

derive the condition satisfied by w^* (the "normal equations").

8) **MAP with Gaussian prior**
Assume

$$\text{Likelihood:} \qquad p(y \mid X, w) \propto \exp\Big(-\frac{1}{2\sigma^2}\|Xw - y\|^2\Big),$$

$$\text{Prior:} \qquad p(w) \propto \exp\Big(-\frac{1}{2}\|w\|^2\Big).$$

Show that maximising the posterior is equivalent to minimising $\|Xw - y\|^2 + \sigma^2\|w\|^2$.

9) **Closed-form solution of the MAP estimator.**
Starting from the MAP objective

$$J_{\text{MAP}}(w) = \|Xw - y\|^2 + \sigma^2\|w\|^2,$$

derive w^*_{MAP}.

10) **Comparing ML and MAP solutions.**
Write down the ML and MAP solutions side by side. How are they related?