Course on probabilistic graphical models
Master MVA
Practice exercises 4

These exercises are not meant to provide an exhaustive coverage of the material to review for the final exam. To some extend they focus more specifically on material that is not covered in the homeworks. Also, all these exercises should not be taken as representative of the difficulty of the questions posed at the exam, although several questions of the exam are likely to have a similar style. Some exercises are easy, and a few can be much harder so don't be discouraged if you find some of them difficult. They are primarily designed to help you review and consolidate your understanding of the course.

Bayesian regression

Consider the Gaussian probabilistic conditional model seen in class for linear regression in which given a pair of variables (X, Y) with X taking values in $\mathbb{R}^d$ and Y in $\mathbb{R}$, we model the conditional distribution of Y given $X = x$ by a Gaussian distribution $\mathcal{N}(w^\top x, \sigma^2)$ parametrized by w and σ^2. Assume that σ^2 is fixed and w unknown and that the problem of learning the linear regression is approached from a Bayesian point of view, by placing a Gaussian prior distribution on w of the form $\mathcal{N}(0, \tau^2 I_d)$.

1. Compute the parameters of the joint distribution of (w, Y).
2. Compute de posterior distribution on w.
3. Compute the predictive distribution over a new output variable y' given a new input x'.

Linear regression and Gaussian likelihood

(a) Let $\psi : H \to \Theta$ be a surjective mapping from $H \in \mathbb{R}^p$ to $\Theta \in \mathbb{R}^p$. Consider the two statistical models

$$\mathcal{P} := \{p_{\psi(\eta)} \mid \eta \in H\}, \quad \text{and} \quad \mathcal{P}' := \{p_\theta \mid \theta \in \Theta\},$$

with $p_\theta = p_{\psi(\eta)}$ when $\theta = \psi(\eta)$. Assume that based on a sample $\mathcal{S}$, the maximum likelihood estimator $\hat{\eta}$ for η in $\mathcal{P}$ exists and is unique. Show that the maximum likelihood estimator $\hat{\theta}$ for θ in $\mathcal{P}'$ exists, is unique, and that $\hat{\theta} = \psi(\hat{\eta})$.

(b) Consider a pair of random variables (X, Y) with $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}$ and with some joint distribution P. Assume that an i.i.d. sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ drawn from P is available. Assume that P is sufficiently nice and n sufficiently large, so that the maximum likelihood estimator for the parameters of the Gaussian model exists and is unique. Let

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_x \\ \hat{\mu}_y \end{pmatrix} \quad \text{and} \quad \widehat{\Sigma} = \begin{pmatrix} \widehat{\Sigma}_{xx} & \widehat{\Sigma}_{xy} \\ \widehat{\Sigma}_{yx} & \widehat{\sigma}^2_y \end{pmatrix}$$

be respectively the maximum likelihood estimator of the mean and covariance matrix with $\hat{\mu}_x \in \mathbb{R}^p$, $\hat{\mu}_y \in \mathbb{R}$, $\widehat{\Sigma}_{xx} \in \mathbb{R}^{p \times p}$, $\widehat{\Sigma}_{xy} = \widehat{\Sigma}_{yx}^\top \in \mathbb{R}^p$, and $\widehat{\sigma}_y^2 \in \mathbb{R}_+$. In particular we assume that $\widehat{\Sigma}$ is invertible.

Assume now that still based on the same sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, a linear regression model of the form $y = w^\top x + b$ is estimated using ordinary least-squares regression (where w and b are the parameters for the uncentered and unnormalized data). Let $\hat{w}$ and $\hat{b}$ be obtained estimators and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{w}^\top x_i - \hat{b})^2$ the average squared residuals. Show that $\hat{w}$, $\hat{b}$ and $\hat{\sigma}^2$ can be expressed as a function of $\hat{\mu}_x, \hat{\mu}_y, \widehat{\Sigma}_{xx}, \widehat{\Sigma}_{xy}$ and $\widehat{\sigma}_y^2$. Explain why and provide the formulas.

Gaussian Markov Chain

The following exercise does not require to do any tedious calculations. If you get into complicated calculation, it means you have the wrong approach...

Assume that the ε_i are i.i.d. with $\varepsilon_i \sim \mathcal{N}(0,1)$ and let $X_1 = \varepsilon_1,\ X_2 = \rho X_1 + \varepsilon_2,\ X_3 = \rho X_2 + \varepsilon_3$.

(a) What is the precision matrix of the joint distribution of (X_1, X_2, X_3)?

(b) Compute $\mathbb{E}[X_2 \mid X_1, X_3]$ and $\mathrm{Var}(X_2 \mid X_1, X_3)$.