1. **Logistic Regression** - Recall the Bayesian logistic regression model from the lectures. We have a labelled dataset $(X, \mathbf{y})$, with $\mathbf{x_i} \in \mathbb{R}^F$ and $y_i \in \{1, -1\}$. Given regression weights $\mathbf{w} \in \mathbb{R}^F$, the likelihood of one data point $(\mathbf{x}, y)$ is

$$p(y|\mathbf{x}, \mathbf{w}) = \begin{cases} \sigma(\mathbf{x}^\top \mathbf{w}) & \text{if } y = 1 \\ 1 - \sigma(\mathbf{x}^\top \mathbf{w}) & \text{if } y = -1 \end{cases} \tag{1}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function. We defined a Gaussian prior distribution for the weights, $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; \mathbf{0}, \sigma_w^2 I)$.

   (a) What is the log-posterior, $\log p(\mathbf{w}|\mathbf{y}, X)$? You may ignore terms that are independent of $\mathbf{w}$ in the expression.

   (b) Calculate the derivative of the log-posterior with respect to $\mathbf{w}$, $\nabla_\mathbf{w} \log p(\mathbf{w}|\mathbf{y}, X)$.

   For both parts, you may find it useful to introduce the notation $\bar{y}$, where $\bar{y} = 1$ when $y = 1$, and $\bar{y} = 0$ when $y = -1$.

2. **Theory Question** - A useful property of exponential family distributions defined in lectures is the existence of *conjugate priors*. Given an exponential family distribution as the likelihood model, if the prior distribution over the parameters has a specific form (the conjugate prior), then the posterior will also be of the same form. The Gaussian distribution is a special case, where the conjugate prior is also a Gaussian. In general, however, the conjugate prior is distinct from the exponential family likelihood.

   You are given count-based data, e.g., a series of spike counts x_i recorded from a neuron, $X = \{x_1, \ldots, x_n\}$, each for a given time window in independent trials. You decide to build an exponential family model, assuming a Poisson likelihood with rate parameter λ, and taking the corresponding conjugate Gamma prior over the rate:

$$\lambda \sim \text{Gamma}(\lambda|k, \theta) = \frac{1}{\Gamma(k)\theta^k} \lambda^{k-1} \, e^{-\frac{\lambda}{\theta}}, \text{ with } \mathbb{E}(\lambda|k, \theta) = k\theta$$
$$x \sim \text{Poisson}(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!},$$

   where k and θ are the fixed shape and scale parameters of the Gamma prior respectively, and Γ is the Gamma function.

   (a) The posterior will be Gamma-distributed—derive the parameters and the mean of the posterior distribution over the firing rate λ, $p(\lambda|X)$.

   (b) Show that the MLE (maximum likelihood estimate) for λ is given by $\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$, and that, for large n, the MLE and the posterior mean coincide.

   (c) Calculate the Laplace approximation to the posterior $p(\lambda|X)$.

   (d) (Optional Challenge) Show that the variance of the Laplace approximation is asymptotically correct (i.e. as $n \to \infty$). The variance of a Gamma distribution is $\text{Var}[\lambda|k, \theta] = k\theta^2$.

3. **Coding Question** - See `Exercise_05.ipynb`