1. Consider the $ETS(AAN)$ model with $\alpha = 0.1$, $\beta = 0.1$, $\ell_{99} = 8$, $b_{99} = 1$, $y_{99} = 10$, $y_{100} = 8$, $\sigma^2 = 16$.

$$\begin{cases} u_t \sim \mathcal{N}(0, \sigma^2) \\ b_t = b_{t-1} + \beta u_t \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t \\ y_t = \ell_{t-1} + b_{t-1} + u_t \end{cases}$$

   (a) Find ℓ_{100}, b_{100}.
   (b) Find 95% predictive interval for y_{102}.

2. [10] I have two exactly proportional regressors, $w_i = 2x_i$, Consider the ridge regression loss function,

$$\text{loss}(\hat{\beta}_1, \hat{\beta}_2) = \sum_i (y_i - \hat{y}_i)^2 + \lambda \cdot (\hat{\beta}_1^2 + \hat{\beta}_2^2), \quad \hat{y}_i = \hat{\beta}_1 x_i + \hat{\beta}_2 w_i.$$

   (a) [4] Write the first order conditions for $\hat{\beta}_1$ and $\hat{\beta}_2$.
   (b) [6] Find the penalized estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for arbitrary $\lambda > 0$.

3. Let's use $\{-1, +1\}$ encoding for binary variable y_i instead of $\{0, 1\}$. Consider a simple logit model where $\mathbb{P}(y_i = 1 \mid x_i) = \Lambda(\beta_1 + \beta_x x_i)$ where x_i is scalar.

   (a) [6] Write the first order conditions for optimization. Explicitely find $\Lambda'(u)$ as a function of u.

   You have 500 observations, $y_i = 1$ in 300 observations.

   (b) [4] Without finding $\hat{\beta}_1$, $\hat{\beta}_2$ find the sum $\sum \hat{\mathbb{P}}(y_i = 1 \mid x_i) = \sum \Lambda(\hat{\beta}_1 + \hat{\beta}_x x_i)$.

4. [10] Consider the model $y = X\beta + u$ where β is non-random, $\mathbb{E}(u \mid X) = 0$. The matrix X of size $n \times k$ has rank $X = k$ and $\mathbb{V}\text{ar}(u \mid X) = \sigma^2 I$. Let $\hat{\beta}$ be the standard OLS estimator of β.

   (a) [2] Find $\mathbb{E}(y - \hat{y} \mid X)$.
   (b) [4 + 4] Find $\mathbb{C}\text{ov}(\hat{\beta}, y \mid X)$ and $\mathbb{C}\text{ov}(\hat{y}, y - \hat{y} \mid X)$.

5. [10] Consider the matrix $X = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$.

   (a) [4] Find the matrix $X^T X$ and diagonalize it.
   (b) [3] Find the SVD of X.

   The generalized Moore-Penrose inverse for the matrix $M = UDV^T$ is defined as $M^+ = VD^+U^T$ where to calculate D^+ we invert all non-zero elements on the diagonal leaving all zeros in place.

   (c) [3] Find the generalized inverse X^+.

6. [10] The prior belief for the parameter x is uniform on $[0; 1]$. Conditionally on x the observations y_1 and y_2 are independent exponentially distributed with rate x.

   (a) [7] Find the posterior distribution of the parameter x given that $y_1 = 5$ and $y_2 = 7$ up to a normalizing constant.
   (b) [3] If possible find the normalizing constant.

   Hint: if you missed the last lecture, then just replace 'posterior' by 'conditional' and 'prior' by 'unconditional' and recall that $f(x \mid y) = f(x, y)/f(y)$.