Homework 4: Due December 20

Reading: Read section 3.3 and 3.4 of course notes.

1. Prove Theorem 3.13.

   Hint: Adapt the proof of Theorem 3.12.

   Proof outline. The liklihood estimator is given by the same formula as in the proof of Theorem 3.12. So when optimizing, again the first term is independent of θ. The second term also has the derivative given in terms of the feature matrix as in the proof of Theorem 3.11. So putting together the two proofs we get the proof of this theorem. The key is to notice that when taking the derivative with respect to θ, it does not matter if we have x_i or $\phi(x_i)$ as the coefficients. □

2. Reformulate and prove Theorem 3.8 for linear models $\mathbb{R}^D \to \mathbb{R}^N$ and the quadratic loss $\ell(y, \hat{y}) = \|y - \hat{y}\|^2$.

   Outline of proof strategy. The difference here is that we now have $y_i \in \mathbb{R}^N$, so the matrix $Y = [y_1 \cdots y_n]^T \in \mathbb{R}^{n\times N}$.

   The outline is provided after Proposition 3.18 in the lecture notes. In particular for the linear model we have

$$\phi(x) = \begin{pmatrix} 1 \\ x \end{pmatrix} \in \mathbb{R}^{D^+1}$$

   So given $\theta = [\theta^{(1)} \cdots \theta^{(N)}] \in \mathbb{R}^{(D+1)\times N}$, with each $\theta^{(i)} = (a^{(i)}, b^{(i)}) \in \mathbb{R} \times \mathbb{R}^D$,

$$f_\theta : \mathbb{R}^D \to \mathbb{R}^N \text{ is given by } f_\theta(x) = \begin{bmatrix} \phi(x)^T\theta^{(1)} \\ \vdots \\ \phi(x)^T\theta^{(N)} \end{bmatrix} = \begin{bmatrix} a^{(1)} + \langle b^{(1)}, x\rangle \\ \vdots \\ a^{(N)} + \langle b^{(N)}, x\rangle \end{bmatrix}$$

   The proof is finished by applying the previous linear case applied to each entry separately. □

3. Prove Proposition 3.14.

   Outline. See the explanation for problem 1 above. Once you have done this problem, the point here is that the first term in the likelihood estimator is now also a variable. Taking the derivative with respect to θ and the derivative with respect to σ gives two equations to set equal to zero. Follow this computation, and see how θ is the same as in Theorem 3.13, and the value for σ also comes from this formula. □

4. From the `RDatasets` package in `Julia` load the `pressure` data set. This data set contains the variables `temperature` and `pressure`, which give the values of pressure of mercury depending on temperature.

   The Antoine equation is a simple model for this dependency: $\log(\texttt{pressure}) = a - \frac{b}{\texttt{temperature}}$.

   Set up and solve a regression problem to estimate a and b.

   See Juila Exercises Homework4_Problem4

5. (**Hand in**) Let $w_1, \ldots, w_n \in \mathbb{R}$. Prove that the median of the w_i minimizes the aggregated distances $d(v) = \sum_{i=1}^n |w_i - v|$. Recall this is used in finding the linear constant when working with Dual SVM.

   Proof idea. We take the derivative of $d(v)$. Notice that the derivative is $+1$ if $w_i - v > 0$ and the derivative if -1 if $w_i - v < 0$. The derivative is not defined when $v = w_i$, but we will set it equal to zero if $w_i = v$ since that term of the sum doesn't actually contribute to the sum. This is the sign function $\text{sign}(w_i - v)$

   Thus setting the derivative equal to zero means

$$\frac{\partial d(v)}{\partial v} = \sum_{i=1}^n \text{sign}(w_i - v) = 0$$

   So we have a sum of $+1, -1, 0$ adds to zero, which means the number of $+1$ and -1 must be the same. If n is odd, then the median of w_i is $w_{(n+1)/2}$, which doesn't guarantee a zero derivative, but still makes the smallest possible value. For example when consdiering $(0, 0, 1)$. When taking an even number the median is $\frac{w_{n/2}+w_{n/2+1}}{2}$. Argue similarly for minimizing the distance. □

6. Exercise 3.8 of the course notes. The `MNIST` database inside `MLDatasets.jl` consists of handwritten digits from 0 to 9. Adapting the `classification` notebook (Notebook 5) and restricting the dataset to pictures of 0s and 1s only, set up a support vector machine to distinguish 0s from 1s.

   Note: when loading the `MNIST` database for the first time, digit "y" when prompted to download the database.

   Follow the outline as given in Notebook 5