Problem Set 2

June 5, 2022

DUE: Sunday June 12 @ midnight EST

1. **Written Problems (50pts)**

   (a) (15pts) Bishop 3.3. Consider a dataset where each data point t_n is associated with a weighting factor $r_n > 0$, so that the sum-of-squares error function becomes

   $$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}^2$$

   (i) Maximum Likelihood Estimator
   (ii) Regularized Least Squares

   Find an expression for the solution $\mathbf{w}^*$ that minimizes this error function. Give two alternative interpretations of the weighted sum-of-squares error function in terms of (i) data dependent noise variance and (ii) replicated data points.

   (b) (15pts) Bishop 3.4. Consider a linear model of the form

   $$y(x, \mathbf{w}) = w_0 + \sum_{i=1}^{D} w_i x_i$$

   together with a sum-of-squares error function of the form

   $$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2$$

   Now suppose that Gaussian noise ϵ_i with zero mean and variance σ^2 is added independently to eaqch of the input variables $\mathbf{x}_i$. By making use of $\mathbb{E}[\epsilon_i] = 0$ and $\mathbb{E}[\epsilon_i\epsilon_j] = \delta_{ij}\sigma^2$, show that minimizing E_D averaged over the noise distribution is equivalent to minimizing the sum-of-squares error for noise-free input variables with the addition of a weight-decay regularizatiopn term, in which the bias parameter w_0 is omitted from the regularizer.

   (c) (20pts) Consider a linear model with the same form and error function as the previous problem:

i. (5 points) Derive the vectorized analyitical solution (also called the *closed form solution*):

$$\mathbf{w}^* = \begin{bmatrix}\mathbf{D} & \mathbf{1}\end{bmatrix}^{\dagger} \mathbf{t}$$

ii. (5 points) Derive the gradient of a single parameter:

$$\frac{\partial E_D(\mathbf{w})}{\partial w_j} = \sum_{n=1}^{N} \{y(\mathbf{x}_n, \mathbf{w}) - t_n\} x_{nj}$$

iii. (10 points) Vectorize the gradient to get the following formula:

$$\nabla_{\mathbf{w}} E_D(\mathbf{w}) = \frac{\partial E_D(\mathbf{w})}{\partial \mathbf{w}} = \begin{bmatrix}\mathbf{D} & \mathbf{1}\end{bmatrix}^{T} \begin{bmatrix} y(\mathbf{x}_1, \mathbf{w}) - t_1 \\ y(\mathbf{x}_2, \mathbf{w}) - t_2 \\ \cdots \\ y(\mathbf{x}_n, \mathbf{w}) - t_n \end{bmatrix}$$

2. **Programming Problem (50pts)** Please go to the following link and make a colab (`https://colab.research.google.com/drive/11QoAtxSEfhRO9Rbb5wyADIpLWWkwRMji?usp=sharing`) and create a copy for youself. This assignment is focused on implementing Linear Regression. You will need the vectorized formulae for the closed-form solution as well as the vectorized form of the gradient from question 1(c). Your job is to complete the `LinearRegression` class in the colab file. When working correctly, the cells after the `LinearRegression` class should run. I have removed my solution from the methods you need to complete, but the output for the cells is still there to give you an idea of what it should look like when your solution works.