ILLINOIS

IE534/CS547: Deep Learning **(Due: Feb-03-2021)**

Homework #1

Instructor: Richard B. Sowers

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- **This is a group homework, every group only submit ONE solution on Compass .** Please include the names of all the group members.
- **Due time is at 11:59pm** at the due date. **No late submission!**
- All students are expected to abide by the Honor Code
- All date-times will be in Champaign-Urbana

Problem 1: gradientdescent

(10 points)

Consider the cost function

$$f(x_1,x_2) \stackrel{\text{def}}{=} 9x_1^2 + x_2^2 \qquad \mathbf{x} = (x_1,x_2) \in \mathbb{R}^2$$

(1) Compute $\nabla f(\mathbf{x})$
(2) Fix $\delta > 0$ and construct gradient descent $\mathbf{x}_{n+1}^{(\delta)} = \mathbf{x}_n^{(\delta)} - \delta\nabla f\left(\mathbf{x}_n^{(\delta)}\right)$, with initial conditions $\mathbf{x}_0^{(\delta)} = (1,2)$. Defining $X_t^{(\delta)} \stackrel{\text{def}}{=} \mathbf{x}_{\lfloor t/\delta \rfloor}^{(\delta)}$ (with $\lfloor\cdot\rfloor$ the integer floor function), find an ODE for $\lim_{\delta\searrow 0} X_t^{(\delta)}$
(3) Explicitly solve this ODE

(Solution)

(1)

$$\nabla f(x_1,x_2) = (18x_1, 2x_2)$$

(2) Since

$$x_{\lfloor \frac{t}{\delta}\rfloor+1} - x_{\lfloor \frac{t}{\delta}\rfloor} = X_{t+\delta} - X_t$$

Then

$$\dot{X}_t = -\nabla f(X_t) = -2AX_t$$

where

$$A \stackrel{\text{def}}{=} \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$$

(3) Given $\mathbf{x}_0^{(\delta)} = (1,2)$

$$X_t = (e^{-18t}, 2e^{-2t})$$

Problem 2: stability

(10 points)

Consider the function

$$f(x) \stackrel{\text{def}}{=} \lambda x^2. \qquad x \in \mathbb{R}$$

where $\lambda > 0$.

(1) Explicitly describe the gradient descent iteration $x_{n+1} = x_n - \delta f'(x_n)$.
(2) Describe the stability of gradient descent iteration for different values of δ.

(Solution)

(1) Since $f'(x) = 2\lambda x$ and $x_{n+1} - x_n - \delta f'(x_n) = x_n - 2\delta\lambda x_n = (1 - 2\delta\lambda)x_n$, thus

$$x_{n+1} = (1 - 2\delta\lambda)x_n.$$

(2) if $|1 - 2\delta\lambda| < 1$, (i.e., $\delta < 1/\lambda$), gradient descent decreases f. If $\delta = 1/\lambda$, gradient descent oscillates, and if $\delta > 1/\lambda$, gradient descent increases f.

Problem 3: Coding question (10 points)

Let's understand linear regression for all features in the dataset used in the linear regression lecture; i.e., let's understand multidimensional linear regression

- The data set is QSAR fish toxicity Data Set and you can also download it on Piazza.
- Attribute information
  1. CIC0
  2. SM1_Dz(Z)
  3. GATS1i
  4. NdsCH
  5. NdssC
  6. MLOGP
  7. quantitative response, LC50 [-LOG(mol/L)]
- The linear regression model is given by $\mathrm{LC50} = \alpha_1\mathrm{CIC0} + \alpha_2\mathrm{SM1\_Dz(Z)} + \alpha_3\mathrm{GATS1i} + \alpha_4\mathrm{MLOGP} + \beta$

(1) Use sklearn to find the formula (i.e., coefficients) for the linear regression
(2) Derive the explicit formula for multidimensional linear regression and implement it in numpy to get explicit coefficients
(3) Construct a gradient descent method for linear regression with the data, and use numpy to implement it to find the fixed point.
(Solution)

(1) Code is attached

(2) Let's assume the linear model is given by $y = Xb + \epsilon$, where the bias term is hidden in X. The derivation is based on least squares estimation, and we need to do the following minimization.

$$\begin{aligned}\min \epsilon^T \epsilon &= (y - Xb)^T (y - Xb) \\ &= y^y - 2b^T X^T y + b^T X^T X b\end{aligned}$$

Let the gradient to be 0, we have

$$\frac{\partial(\epsilon^T \epsilon)}{\partial b} = -2X^T y + 2X^T X b = 0$$

Then we have

$$b = (X^T X)^{-1} X^T y$$

(3) Code is attached