CS5020: Nonlinear Optimisation: Theory and Algorithms
Coding Exercise - 3 (5 Marks)

Linear Regression

(1) Choose your favourite slope $m \in \mathbb{R}$ and intercept $c \in \mathbb{R}$ (pick m, c in the range $(-5, 5)$). For this problem $x_* = (m, c)$. Generate $n = 100$ datapoints $(a_i, b_i)_{i=1}^n$ such that

$$b_i = ma_i + c + \text{Uniform}_i(-1, 1)$$

,where $\text{Uniform}_i(-1, 1), i = 1, \dots, 1000$ are i.i.d uniformly distributed in the range $(-1, 1)$. The loss function is given by

$$f(x(1), x(2)) = \frac{1}{n}\sum_{i=1}^{n} f_i(x(1), x(2)) = \frac{1}{n}\sum_{i=1}^{n} (b_i - (a_i x(1) + x(2)))^2$$

(i) Run gradient descent for $t = 1, \dots, T$ to find $x(1), x(2)$. Try out values such as $\alpha = 0.01, 0.1$ and $T = 100, 1000$ to fix the value of T that achieves good convergence.

(ii) Use stochastic gradient descent to find $x(1), x(2)$, by choosing sample size $S = 10$ and gradient as below. Try out values such as $\alpha = 0.01, 0.1$ and $T = 100, 1000$ to fix the value of T that achieves good convergence.

$$\widehat{\nabla} f(x(1), x(2)) = \frac{1}{S}\sum_{i \in S} \nabla f_i(x(1), x(2))$$

(a) Set $\alpha = 0.01$, plot $||x_t - x_*||_2$ versus t for gradient descent and stochastic gradient descent in the same figure.

(b) Set $\alpha = 0.1$, plot $||x_t - x_*||_2$ versus t for gradient descent and stochastic gradient descent in the same figure.

(c) Set $\alpha = 0.1$, $T = 1000$. Plot the data in the dataset, and in the same figure, plot the fit $b \approx ax_t(1) + x_t(2)$ as a function of $t = 1, \dots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation. Your output should be like Template 1 in the shared python notebook.

(d) Set $\alpha = 0.01$, $T = 1000$. Plot the data in the dataset, and in the same figure, plot the fit $b \approx ax_t(1) + x_t(2)$ as a function of $t = 1, \dots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation. Your output should be like Template 1 in the shared python notebook.

(e) Set $\alpha = 0.01$, $T = 1000$. In the same figure, make a 3D-plot $f(x)$ and mark points $f(x_t)$ as a function of $t = 1, \dots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation. Your output should be like Template 2 in the shared python notebook.

(f) Set $\alpha = 0.01$, $T = 1000$. In the same figure, make a 3D-plot $f(x)$ and mark points $f(x_t)$ as a function of $t = 1, \dots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation. Your output should be like Template 2 in the shared python notebook.

(g) Set $\alpha = 0.01, T = 1000$. In the same figure, quiver plot the negative of the gradients and mark points $(x_t(1), x_t(2))$ as a function of $t = 1, \dots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation. Your output should be like Template 3 in the shared python notebook.

(h) Set $\alpha = 0.01$, $T = 1000$. In the same figure, quiver plot the negative of the gradients and mark points $(x_t(1), x_t(2))$ as a function of $t = 1, \dots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation. Your output should be like Template 3 in the shared python notebook.

Logistic Regression

(2) Generate a dataset $(a_i, b_i)_{i=1}^n$ for $n = 100$ as follows.

• Let $b_i = -1$ for $i = 1, \ldots, 50$ denote the group kids and $b_i = +1$ for $i = 51, \ldots, 100$ denote group adults.

• Each $a_i = (\text{weight}_i, \text{height}_i)$ of the i^{th} individual.

• For $i = 1, \ldots, 50$ sample the weight from Uniform$(30, 45)$ and sample the height from Uniform$(125, 145)$.

• For $i = 51, \ldots, 100$ sample the weight from Uniform$(55, 70)$ and sample the height from Uniform$(155, 180)$.

Learn the line $x(1)a(1) + x(2)a(2) + x(3) = 0$ that separates the kids from the adults by minimising the following function

$$\begin{aligned} f(x(1), x(2), x(3)) &= \frac{1}{n}\sum_{i=1}^{n} f_i(x(1), x(2), x(3)) \\ &= \frac{1}{n}\sum_{i=1}^{n} \ln\left(1 + \exp\left(-b_i\Big(x(1)a_i(1) + x(2)a_i(2) + x(3)\Big)\right)\right) \end{aligned}$$

(i) Run gradient descent for $t = 1, \ldots, T$ to find $x(1), x(2), x(3)$. Try out values such as $\alpha = 0.01, 0.1$ and $T = 100, 1000$ to fix the value of T that achieves good convergence.

(ii) Use stochastic gradient descent to find $x(1), x(2), x(3)$, by choosing sample size $S = 10$ and gradient as below. Try out values such as $\alpha = 0.01, 0.1$ and $T = 100, 1000$ to fix the value of T that achieves good convergence.

$$\widehat{\nabla} f(x(1), x(2)) = \frac{1}{S}\sum_{i \in S} \nabla f_i(x(1), x(2))$$

(a) Set $\alpha = 0.1$, $T = 1000$. Plot the data in the dataset, and in the same figure, plot the line $a(1)x_t(1) + a(2)x_t(2) + x(3) = 0$ as a function of $t = 1, \ldots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation.

(b) Set $\alpha = 0.01$, $T = 1000$. Plot the data in the dataset, and in the same figure, plot the line $a(1)x_t(1) + a(2)x_t(2) + x(3) = 0$ as a function of $t = 1, \ldots, T$ for gradient descent and stochastic gradient descent. Use matplotlib animation.