M124 Coursework 1

Vangelis Tsiatouras
cs2190021@di.uoa.gr

March 25, 2021

1 Vector and Matrices

The generated matrices using NumPy are the following:

$$X = \begin{bmatrix} -13 & -38 & 22 & -41 \\ 25 & -45 & 29 & 14 \\ -34 & -49 & 26 & 21 \end{bmatrix}, \quad Y = \begin{bmatrix} -44 & -25 & 0 \\ -30 & -32 & 34 \\ -39 & -22 & -21 \\ -36 & 0 & 18 \end{bmatrix},$$

$$g = \begin{bmatrix} 37 & 37 & 44 & 46 \end{bmatrix}, \quad z = \begin{bmatrix} 36 & -37 & -41 & -43 \end{bmatrix}$$

- Compute the inner product of g and $z \in \mathbb{Z}$.

  $$\langle g, z \rangle = 37 \cdot 36 + 37 \cdot (-37) + 44 \cdot (-41) + 46 \cdot (-43) = 1332 - 1369 - 1804 - 1978 = -3819$$

- Compute the inner product of g and $g^T z \in \mathbb{Z}$.

  Initially, $g^T z$ is:

  $$\begin{aligned} g^T z &= \begin{bmatrix} 37 \\ 37 \\ 44 \\ 46 \end{bmatrix} \cdot \begin{bmatrix} 36 & -37 & -41 & -43 \end{bmatrix} \\ &= \begin{bmatrix} 37 \cdot 36 & 37 \cdot (-36) & 44 \cdot (-41) & 46 \cdot (-43) \end{bmatrix} \\ &= \begin{bmatrix} 1332 & -1369 & -1804 & -1978 \end{bmatrix} \end{aligned}$$

  Then, the inner product is:

  $$\langle g, g^T z \rangle = 37 \cdot 1332 + 37 \cdot (-1369) + 44 \cdot (-1804) + 46 \cdot (-1978) = 49284 - 50653 - 79376 - 90988 = -1717333$$

- Compute the matrix-vector product $Xg \in \mathbb{Z}^3$.

$$
\begin{aligned}
Xg &= \begin{bmatrix} -13 & -38 & 22 & -41 \\ 25 & -45 & 29 & 14 \\ -34 & -49 & 26 & 21 \end{bmatrix} \cdot \begin{bmatrix} 36 \\ -37 \\ -41 \\ -43 \end{bmatrix} \\
&= \begin{bmatrix} -13 \cdot 37 - 38 \cdot 37 + 22 \cdot 44 - 41 \cdot 46 \\ 25 \cdot 37 - 45 \cdot 37 + 29 \cdot 44 + 14 \cdot 46 \\ -34 \cdot 37 - 49 \cdot 37 + 26 \cdot 44 + 21 \cdot 46 \end{bmatrix} \\
&= \begin{bmatrix} -481 - 1406 + 22 + 968 - 1886 \\ 925 - 1665 + 1276 + 644 \\ -1258 - 1813 + 1144 + 966 \end{bmatrix} \\
&= \begin{bmatrix} -2805 \\ 1180 \\ -961 \end{bmatrix}
\end{aligned}
$$

- Compute the dot product $XY \in \mathbb{Z}^{3\times 3}$.

$$
\begin{aligned}
XY &= \begin{bmatrix} -13 & -38 & 22 & -41 \\ 25 & -45 & 29 & 14 \\ -34 & -49 & 26 & 21 \end{bmatrix} \cdot \begin{bmatrix} -44 & -25 & 0 \\ -30 & -32 & 34 \\ -39 & -22 & -21 \\ -36 & 0 & 18 \end{bmatrix} \\
&= \begin{bmatrix} 572 + 1140 - 858 + 1476 & 325 + 1216 - 484 + 0 & 0 - 1292 - 462 - 738 \\ -1100 + 1350 - 1131 - 504 & -625 + 1440 - 638 + 0 & 0 - 1530 - 609 + 252 \\ 1496 + 1470 - 1014 - 756 & 850 + 1568 - 572 + 0 & 0 - 1666 - 546 + 378 \end{bmatrix} \\
&= \begin{bmatrix} 2330 & 1057 & -2492 \\ -1385 & 177 & -1887 \\ 1196 & 1846 & -1834 \end{bmatrix}
\end{aligned}
$$

- Compute the L_2 /Frobenius norms of the vectors and matrices.

$$
\begin{aligned}
\|X\| &= \sqrt{\sum_{i,j} x_{i,j}^2} \\
&= \sqrt{(-13)^2 + (-38)^2 + 22^2 + (-41)^2 + 25^2 + (-45)^2 + 29^2 + 14^2 + (-34)^2 + (-49)^2 + 26^2 + 21^2} \\
&= 110.17713011328621
\end{aligned}
$$

$$
\begin{aligned}
\|Y\| &= \sqrt{\sum_{i,j} x_{i,j}^2} \\
&= \sqrt{(-44)^2 + (-25)^2 + (-30)^2 + (-32)^2 + 34^2 + (-39)^2 + (-22)^2 + (-21)^2 + (-36)^2 + 18^2} \\
&= 98.52410872471773
\end{aligned}
$$

$$
\|g\| = \sqrt{\sum_{i} x_i^2} = \sqrt{37^2 + 37^2 + 44^2 + 46^2} = 82.40145629781065
$$

$$
\|z\| = \sqrt{\sum_{i} x_i^2} = \sqrt{36^2 + (-37)^2 + (-41)^2 + (-43)^2} = 78.70832230456955
$$

2 Gradient Computation

1. It is known that

$$\frac{\partial(\mathbf{x}^T\mathbf{b})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{x}\mathbf{b}^T)}{\partial\mathbf{x}} = \mathbf{b}^T \qquad \textit{(first order)} \tag{1}$$

   Also

$$f(g(\mathbf{x}), h(\mathbf{x})) = \langle g(\mathbf{x}), h(\mathbf{x})\rangle = g^T(\mathbf{x})h(\mathbf{x})$$

   By assigning $g(\mathbf{x}) = \mathbf{x}$ & $h(\mathbf{x}) = \mathbf{A}\mathbf{x}$, f becomes:

$$f(g(\mathbf{x}), h(\mathbf{x})) = \mathbf{x}^T\mathbf{A}\mathbf{x}$$

   By applying the multiplication rule and keeping in mind that $\mathbf{A}$ is symmetrical ($\mathbf{A} = \mathbf{A}^T$), we can yield:

$$\begin{aligned}\frac{df(g(\mathbf{x}), h(\mathbf{x}))}{d\mathbf{x}} &= g^T(\mathbf{x})\frac{\partial h(\mathbf{x})}{\partial\mathbf{x}} + h^T(\mathbf{x})\frac{\partial g(\mathbf{x})}{\partial\mathbf{x}} \\ &= \mathbf{x}^T\frac{\partial\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} + (\mathbf{A}\mathbf{x})^T\frac{\partial\mathbf{x}}{\partial\mathbf{x}} \\ &= \mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T\mathbf{I} \\ &= \mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T \\ &= \mathbf{x}^T(\mathbf{A} + \mathbf{A}^T) \\ &= 2\mathbf{A}^T\mathbf{x}^T\end{aligned} \tag{2}$$

   By combining (1) & (2) the gradient is:

$$\nabla f = \left(\frac{df}{d\mathbf{x}}\right)^T = (2\mathbf{A}^T\mathbf{x}^T + \mathbf{b}^T)^T = 2\mathbf{A}\mathbf{x} + \mathbf{b}$$

2. The derivative of $c(x)$ by applying the chain rule is:

$$c'(x) = f'(g(h(x))) \cdot g'(h(x)) \cdot h'(x)$$

   By replacing $f(x) = sin(x)$, $g(x) = x^3$ & $h(x) = 5x$ we extract:

$$\frac{dc(x)}{dx} = cos((5x)^3) \cdot 3(5x)^2 \cdot 5 = 375x^2cos(125x^3)$$

3. The *Hessian matrix* can be yielded through the formula

$$H = \nabla^2 f(\mathbf{x}) = \frac{\partial^2 f(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^T}$$

   Although the partial derivative of f can be presented as:

$$\frac{\partial f(x)}{\partial x_k} = \sum_{j=1}^{d} a_{jk}x_j + \sum_{j=1}^{d} a_{kj}x_j + \sum_{j=1}^{d} b_j$$

   Furthermore, the second partial derivatives have the form:

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_l} = a_{kl} + a_{lk}$$

   This results to

$$H = \nabla^2 f(\mathbf{x}) = \mathbf{A} + \mathbf{A}^T = 2\mathbf{A}$$

3 Linear Regression

Derivation of the Ordinary Least Squares estimator for multiple regressors

In order to derive the Least Squares estimates for the multivariate Linear Regression case we should expand further the linear predictor $f(\mathbf{X}) = \mathbf{b}^T\mathbf{X}$. Given that the vector $\mathbf{b}$ contains the bias factor b_0 and the coefficients $b_1 ... b_n$ and the matrix $\mathbf{X}$ contains all the independent variables, f can be represented as following:

$$f = b_0 + b_1x_1 + b_2x_2 + ... + b_{k-1}x_{k-1} + b_kx_k,$$

where k is the number of features

Also this is derived to the following:

$$f(\mathbf{X}) = \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \textit{(Train target values)}$$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} \qquad \textit{(Bias factor \& coefficients)}$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & ... & X_{1,k} \\ 1 & X_{2,1} & X_{2,2} & ... & X_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,1} & X_{n,2} & ... & X_{n,k} \end{bmatrix} \qquad \textit{(Observations)}$$

where n is the number of the observations

In order to retrieve the beta matrix we should minimize the error:

$$\begin{aligned} Q(\mathbf{b}) &= \|\mathbf{Y} - \mathbf{X}\mathbf{b}\|^2 \\ &= (\mathbf{Y}^T - \mathbf{b}^T\mathbf{X}^T)(\mathbf{Y} - \mathbf{b}\mathbf{X}) \\ &= \mathbf{Y}^T\mathbf{Y} + \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{b} - \mathbf{Y}^T\mathbf{X}\mathbf{b} - \mathbf{b}^T\mathbf{X}^T\mathbf{Y} \\ &= \mathbf{Y}^T\mathbf{Y} + \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{b} - 2\mathbf{b}^T\mathbf{X}^T\mathbf{Y} \end{aligned}$$

$$\frac{\partial Q}{\partial \mathbf{b}} = 0 \Rightarrow 2\mathbf{X}^T\mathbf{X}b - 2\mathbf{Y}\mathbf{X}^T = 0 \Rightarrow \mathbf{X}^T\mathbf{X}b = \mathbf{X}^T\mathbf{Y} \xRightarrow{\cdot(\mathbf{X}^T\mathbf{X})^{-1}} = \mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Concluding, the Least Squares estimates for the multivariate Linear Regression case can be extracted through the formula $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.

The rest questions of part 3 are answered inside the Jupyter Notebook `linear_regression.py`.