Artificial Intelligence and Machine Learning 2023/2024

Week 7 Tutorial and Additional Exercises

Dr Ruchit Agrawal

School of Computer Science

March 4, 2024

In this tutorial. . .

In this tutorial we will be covering

- Univariate and multivariate linear regression.
- Exercises on gradient descent.
- Exercises on geometric concepts.
- Optional theoretical exercises.

Univariate Linear Regression

Recall the formal statement of *univariate linear regression*:

- Given a training set $\{(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})\}$, train weights w_0, w_1 that minimise a loss function.
- Given this training set, and weights w_0, w_1, the *square loss* (or L_2 loss) function is given as

$$g(w_0, w_1) = \sum_{i=1}^{n} (w_0 + w_1 x_{(i)} - y_{(i)})^2.$$

- Informally, we need w_0, w_1 such that for all $i = 1, \ldots, n$

$$w_0 + w_1 x_{(i)} \approx y_{(i)}.$$

Multivariate Linear Regression

Recall the formal statement of *multivariate linear regression*:

- Given a training set $\{(\mathbf{x}_{(1)}, y_{(1)}), \ldots, (\mathbf{x}_{(n)}, y_{(n)})\}$, train a weight vector $\mathbf{w}$ that minimises a loss function.
- If we have d variables, then for all $i = 1, \ldots, n$, we write

$$\mathbf{x}_{(i)} = (1, x^1_{(i)}, x^2_{(i)}, \ldots, x^d_{(i)}) \text{ and } \mathbf{w} = (w_0, w_1, w_2, \ldots, w_d).$$

- Given this training set and a weight vector $\mathbf{w}$, the *square loss* (or L_2 loss) function is given as

$$g(\mathbf{w}) = \sum_{i=1}^{n} (\mathbf{w}^T \mathbf{x}_{(i)} - y_{(i)})^2.$$

- Informally, we need $\mathbf{w}$ such that for all $i = 1, \ldots, n$

$$\mathbf{w}^T \mathbf{x}_{(i)} \approx y_{(i)}.$$

Exercise 1

Consider a univariate linear regression problem with the *mean* square loss:

$$g(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x_{(i)} - y_{(i)})^2$$

- We have this training set of size $n = 4$:

i	$x^{(i)}$	$y^{(i)}$
1	1	3
2	0	2
3	2	5
4	-1	0

Weights w_0, w_1	Loss $g(w_0, w_1)$
$w_0 = 2,\ w_1 = 3$	?
$w_0 = 3,\ w_1 = 1$	?
$w_0 = 2,\ w_1 = 2$	?
$w_0 = 0,\ w_1 = 2$	?

- Fill in the table to the right for each choice of weights.
- Which of these weights yield the minimum loss?

Exercise 1: Solution

- The table is filled as follows:

Weights w_0, w_1	Loss $g(w_0, w_1)$
$w_0 = 2,\ w_1 = 3$	3.5
$w_0 = 3,\ w_1 = 1$	1.5
$w_0 = 2,\ w_1 = 2$	0.5
$w_0 = 0,\ w_1 = 2$	2.5

- The optimal weights out of these are $w_0 = 2, w_1 = 2$.

Exercise 2

Consider the following algorithm.

Algorithm 1: Single iteration of Gradient Descent for Univariate Linear Regression.

Input: Training set: $\{(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})\}$, learning rate: α

Output: Cost C; weights w_0, w_1.

$C \leftarrow 0$;
$w_0 \leftarrow 0$;
$w_1 \leftarrow 0$;
for $i = 1, \ldots, n$ **do**
  $f \leftarrow w_0 + w_1 x_{(i)}$;
  $C \leftarrow C + (f - y_{(i)})^2$;
  $w_0 \leftarrow w_0 - \alpha \cdot (f - y_{(i)})$;
  $w_1 \leftarrow w_1 - \alpha \cdot (f - y_{(i)}) x_{(i)}$.
return C, w_0, w_1.

- What are the numerical values of C, w_0, w_1 at the end of algorithm 1 for $\alpha = 1$ and the following training set of size $n = 3$:

i	$x_{(i)}$	$y_{(i)}$
1	1	1
2	2	5
3	3	11

- For each $i = 0, 1, 2, 3$, we write the values of C, w_0, w_1:

i	C	w_0	w_1
0	0	0	0
1	1	1	1
2	5	3	5
3	54	-4	-16

- Therefore, at the end of algorithm 1, we will have: $C = 54$, $w_0 = -4$, $w_1 = -16$.
- Draw this table for the same training set and $\alpha = 2$. Then, for $\alpha = 0.5$.

Exercise 3

Consider the following pairs of points in the form (x, y). In each case, find the equation of the line that passes between the two given points in the form $y = ax + b$. Also, find its slope.

1. $(1, 2)$ and $(-1, -4)$.
2. $(-1, 3)$ and $(3, -5)$.
3. $(-2, -3)$ and $(1, 0)$.
4. $(3, 5)$ and $(0, 5)$.

Hint: You should find the values of a and b. The slope equals a.

Exercise 3: Solution

The line equations are (in the same order):

1. $y = 3x - 1$; slope is 3.
2. $y = -2x + 1$; slope is -2.
3. $y = x - 1$; slope is 1.
4. $y = 5$; slope is 0.

Exercise 4

In each case, find the point of intersection of the two given lines.

1. $y = x + 1$ and $y = 4x - 2$.
2. $y = 5x$ and $y = -3x$.
3. $y = -2x + 3$ and $y = 4x - 6$.
4. $y = 5$ and $y = -x - 10$.

Hint: In each case, equate the two right-hand-sides to find x. Then solve for y.

Exercise 4: Solution

The points of intersection are (in the same order):

1. $(1, 2)$.
2. $(0, 0)$.
3. $(1.5, 0)$.
4. $(-15, 5)$.

Optional Material

Optional Exercise 1

- Assume that we have trained a multi-variable regression model such that given an instance $\mathbf{x}$, it predicts its y value to be

$$\mathbf{w}^T\mathbf{x}.$$

- Prove that if the model predicts the same value $\hat{y}$ for instances $\mathbf{x}_1$ and $\mathbf{x}_2$, then, for all t, it also predicts the value $\hat{y}$ for the instance $\mathbf{x}_0$, where

$$\mathbf{x}_0 = t\mathbf{x}_1 + (1-t)\mathbf{x}_2.$$

- Geometrically, $\mathbf{x}_0$ lies in the line that passes from $\mathbf{x}_1$ and $\mathbf{x}_2$.
- Hint: Start with $\mathbf{w}^T\mathbf{x}_0$ and expand $\mathbf{x}_0$ according to its formula.

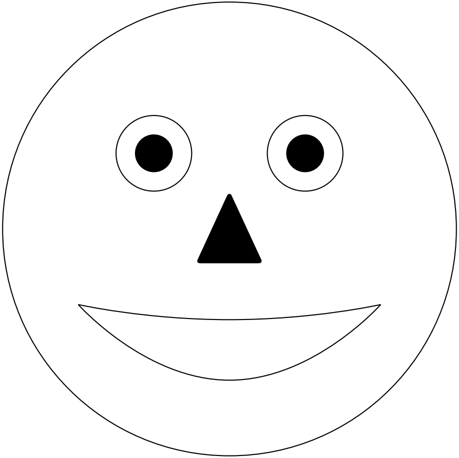
Optional Exercise 1: Solution

- The prediction for $\mathbf{x}_0$ is

$$\begin{aligned} \mathbf{w}^T\mathbf{x}_0 &= \mathbf{w}^T(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \\ &= t\mathbf{w}^T\mathbf{x}_1 + (1-t)\mathbf{w}^T\mathbf{x}_2 \\ &= t\hat{y} + (1-t)\hat{y} \\ &= \hat{y}. \end{aligned}$$

- Therefore the same value $\hat{y}$ is predicted by the model for $\mathbf{x}_0$.

Any questions?



Thank you for your attention!