Artificial Intelligence I 2023/2024

Week 3 Tutorial and Additional Exercises

Linear Regression & Gradient Descent

School of Computer Science

February 13, 2024

In this tutorial. . .

In this tutorial we will be covering

- Univariate and multivariate linear regression.
- Exercises on gradient descent.
- Exercises on geometric concepts.
- Advanced 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

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

- 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

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

- 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 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

Exercise 2: Solution

- 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. $(-15, 5)$.

Advanced Material

(OPTIONAL) Advanced 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

$$\hat{y} := \mathbf{w}^T \mathbf{x}.$$

- If the model predicts the same value $\hat{y}$ for two different instances $\mathbf{x}_1$ and $\mathbf{x}_2$, then, for any real value t, which can generate a new instance $\mathbf{x}_0$, where

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

  prove that $\mathbf{x}_0$ is also predicted as $\hat{y}$ by the model.
- Hint: Start with $\mathbf{w}^T \mathbf{x}_0$ and expand $\mathbf{x}_0$ according to its formula. Geometrically, $\mathbf{x}_0$ lies in the line that passes from $\mathbf{x}_1$ and $\mathbf{x}_2$.

(OPTIONAL) Advanced 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 $\mathbf{x}_0$.

Advanced Exercise 2

- Let $(\mathbf{x}, y)$ be a data point and $\mathbf{w}$ be the weight vector to be optimised in a multivariate linear regression model with d variables. Assume that $\mathbf{x}$ and $\mathbf{w}$ are of the form[1]

$$\mathbf{x} = (x_0, x_1, \ldots, x_d) \text{ and } \mathbf{w} = (w_0, w_1, \ldots, w_d).$$

- Let g be a square loss function of the form

$$g(\mathbf{w}) = (\mathbf{w}^T\mathbf{x} - y)^2.$$

- Use the derivative rules to prove that

$$\nabla g(\mathbf{w}) = 2(\mathbf{w}^T\mathbf{x} - y)\mathbf{x}.$$

- Hint: Find each partial derivative separately, then factor.

[1] We usually take $x_0 = 1$, but we leave it as x_0 here.

Advanced Exercise 2: Solution

- We first write the loss function g as

$$g(w_0, w_1, \ldots, w_d) = (w_0x_0 + w_1x_1 + \cdots + w_dx_d - y)^2.$$

- The partial derivative of g, with respect to w_i, $0 \leq i \leq d$, is

$$\begin{aligned}\frac{\partial g}{\partial w_i}(w_0, w_1, \ldots, w_d) &= 2(w_0x_0 + w_1x_1 + \ldots + w_dx_d - y)x_i \\ &= 2(\mathbf{w}^T\mathbf{x} - y)x_i.\end{aligned}$$

- Therefore, the gradient vector of g is

$$\begin{aligned}\nabla g(\mathbf{w}) &= (2(\mathbf{w}^T\mathbf{x} - y)x_0, 2(\mathbf{w}^T\mathbf{x} - y)x_1, \ldots, 2(\mathbf{w}^T\mathbf{x} - y)x_d) \\ &= 2(\mathbf{w}^T\mathbf{x} - y)(x_0, x_1, \ldots, x_d) \\ &= 2(\mathbf{w}^T\mathbf{x} - y)\mathbf{x}.\end{aligned}$$

(OPTIONAL) Advanced Exercise 3

- A multi-variable function g is called *convex* if and only if for all $\mathbf{w}_1$ and $\mathbf{w}_2$ and for all $0 \leq t \leq 1$ we have

$$g(t\mathbf{w}_1 + (1-t)\mathbf{w}_2) \leq tg(\mathbf{w}_1) + (1-t)g(\mathbf{w}_2).$$

- Convex functions are easy to minimise, and are common choices for loss functions, due to their property that any local minimum is also a global minimum (Try to prove this also!).
- Prove that given a data point $(\mathbf{x}, y)$ and a weight vector $\mathbf{w}$, the following square loss function g is convex:

$$g(\mathbf{w}) = (\mathbf{w}^T\mathbf{x} - y)^2.$$

- Hint: Use the fact that for all real numbers a, b and for all $0 \leq t \leq 1$, we have $(ta + (1-t)b)^2 \leq ta^2 + (1-t)b^2$.

(OPTIONAL) Advanced Exercise 3: Solution

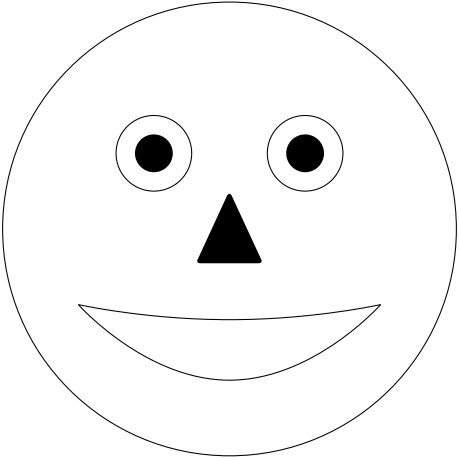
- Let $\mathbf{w}_1, \mathbf{w}_2$ be any weight vectors and $0 \leq t \leq 1$. We have

$$\begin{aligned} g(t\mathbf{w}_1 + (1-t)\mathbf{w}_2) &= ((t\mathbf{w}_1 + (1-t)\mathbf{w}_2)^T\mathbf{x} - y)^2 \\ &= (t\mathbf{w}_1^T\mathbf{x} - ty + (1-t)\mathbf{w}_2^T\mathbf{x} - (1-t)y)^2 \\ &= (t(\mathbf{w}_1^T\mathbf{x} - y) + (1-t)(\mathbf{w}_2^T\mathbf{x} - y))^2 \\ &\leq t(\mathbf{w}_1^T\mathbf{x} - y)^2 + (1-t)(\mathbf{w}_2^T\mathbf{x} - y)^2 \\ &= tg(\mathbf{w}_1) + (1-t)g(\mathbf{w}_2). \end{aligned}$$

- We used the hint to obtain the inequality, since $(\mathbf{w}_1^T\mathbf{x} - y)$ and $(\mathbf{w}_2^T\mathbf{x} - y)$ are real numbers and $0 \leq t \leq 1$.
- Therefore, g is convex.

Any questions?

Until the next time...



Thank you for your attention!