# **Assignment2\_Machine Learning**

1) ما هو الانحدار الخطي؟ هل يمكنك شرح الافكار الأساسية وراء الانحدار الخطي؟ لماذا يُفضل الانحدار على التصنيف؟

الحل:

يعتبر الانحدار الخطي أسلوبًا إحصائيًا يُستخدم لنمذجة العلاقة بين متغير (dependent) ومتغير (independent) او اكتر. مما يعني انه إيجاد أفضل خط ملائم يُمثل العلاقة بين المتغيرات (Best Fit).

## الأفكار الرئيسية وراء الانحدار الخطى تتضمن:

- الخطية (Linearity): يُفترض في الانحدار الخطي أن العلاقة بين المتغيرات هي خطية، أي أنها يمكن تمثيلها بشكل مستقيم.
  - التكيف الأمثل (Optimal Adjustment): الهدف هو العثور على الخط الذي يُمثل بشكل أمثل العلاقة بين المتغيرات بحيث يتنبأ بالقيم المستقبلية بدقة.
- المثالية بواسطة (MSE): في حالة الانحدار الخطي، يتم استخدام أسلوب "أصغر المربعات" لتقدير الخط الأمثل الذي يقلل مجموع مربعات الأخطاء بين البيانات الفعلية والقيم المتوقعة.
  - التحقق والتقييم (Validation and Evaluation): بعد تدريب النموذج، يتم اختباره او تقييم أدائه باستخدام بيانات تحقق غير المستخدمة في التدريب لضمان قدرته على التعامل مع البيانات الجديدة بشكل صحيح.

النموذج او (Model):

y = b + mx

حيث:

y هو المتغير المعتمد (الذي يتم التنبؤ به) هو x المتغير المستقل هو x المتغير المستقل هو m الخط الميل (slope و coefficients) هو b الانقطاع مع محور y

يتم تفضيل (Regression) على (Classification) عندما يكون المخرج الذي نتنبأ به قيمة مستمرة (تنتمي الى IR)، مثل توقع الأسعار أو درجات الحرارة أو الدرجات. بينما يُستخدم التصنيف عندما يكون المخرج منتميًا إلى فئات متميزة، مثل تصنيف رسائل البريد الإلكتروني على أنها رسائل (not spam) او (spam) مثال اخر توقع ما إذا كان المريض مصابًا بمرض أم لا. تعتمد الاختيار بين الانحدار والتصنيف على طبيعة المشكلة ونوع الناتج المرغوب.

## **Pseudocode**

Return theta0, theta1

```
Inputs: x (features), y (labels)
Outputs: theta0 (intercept), theta0 (slope)
Function LinearRegression(x, y):
  Initialize parameters theta with random values or zeros.
  Define learning rate alpha.
  Define number of iterations.
  Loop for i from 1 to iterations:
       Predicted = theta0 + x * theta1
                                                                    // Computing predicted values
       Calculate cost = (1 / 2 * m) * sum((Predicted - y)^2)
                                                                    // Calculating cost function with (MSE)
       derivative_theta0 = (1 / n) * sum(Predicted - y)
                                                                   // Calculate derivative
       derivative_theta1 = (1 / n) * sum((Predicted - y) * x)
       theta0 = theta0 - alpha * derivative_ theta0
                                                                   // Update parameters
       theta 1= theta1 - alpha * derivative_ theta1
```

// Return updated parameters theta0, theta1

إذا استخدمت الانحدار الشريحي (Batch Gradient Descent) ورسمت خطأ التحقق (validation error) في
 كل مرة، والاحظت أن خطأ التحقق يزداد باستمرار، فما الذي يحدث على الأرجح؟ وكيف يمكنك حل هذه المشكلة؟

### الحل:

إذا لاحظت أن خطأ التحقق يزداد باستمرار خلال كل فترة من الانحدار الشريحي، فإن الأمر الأكثر احتمالًا هو حدوث ظاهرة (Overfitting)، وهذا يعني أن النموذج تعلم البيانات التدريبية بشكل زائد وبداء يتعرف على الضوضاء في هذه البيانات، ، مما يؤدي إلى أداء جيد في مجموعة التدريب ولكن بشكل سيئ على البيانات التي لم يراها من قبل مثل(بيانات جديدة او بيانات الاختبار).

## لمعالجة هذه المشاكل نستعمل بعض الطرق التالية:

- تقليص تعقيد النموذج: قد يكون النموذج معقدًا أكثر مما ينبغي. يجب تقليص عدد (Feature) أو استخدام تقليات التنظيم مثل التنظيم (Regularization) أو L2 لتقليات التنظيم مثل التنظيم (Feature)
- زيادة حجم البيانات التدريبية: زيادة حجم البيانات التدريبية قد تقلل من حدوث (Overfitting)عن طريق توفير مجموعة متنوعة من الأمثلة للنموذج ليتعلم منها.
  - ضبط معدل التعلم (α): قم بتقليل معدل التعلم للسماح لعملية التحسين بالتقدم بشكل تدريجي، مما يسمح للنموذج بالتقارب بشكل أكثر سلاسة.
- التقييم المتقاطع (Cross-Validation): استخدم تقنيات التقييم المتقاطع لتقييم أداء النموذج على مجموعات التحقق المتعددة، مما يساعد على تقييم الأداء بشكل أكثر.

		(4
	سليمان حسن المره مانلي 2180103888	
	part one:-	
	(2+3-) (pt) (mon + 0 + 0	10.
	من الحدول التالي: -	14
		- 0
	Hours Studied 0 0.5 1 1 1.5 2 2.75 3	
	Test Score 75 81 78 85 89 80 97 90	
	N = 9  Data distribution	
	~ = 0.0002	
	itreation = 2000	
	ha) = 0, +0, X;	-1-
1=	Q:	
	00 05 10 13 20 25 30	
	ha 00000000	EST
	ha - y 2 5625 6561 6084 7225 7921 6400 9409 8100 57325	
	Semiles - B	Ш,
	MSE 1 2 (how-v)	
	(13E) = 2h [ 1 C 100 ]	
	57325 2500 01	-0.5
	= 5+325 = 35821811	
	6 - 75 40840p	-
	Data dati tu dare	
-4	200-	
	3 mm g 50 mm g	
	200 - 200 -	
	Nation of Health	

	Mark Market No. 1985 Conference of the Conferenc	
	B = B - ~ ((h) ≥ (h∞-y)	
	0 0 - ((1) ((3-)	
TOWN TO	0 0-0.0902 ((g) (-675) 0=0.016875	
TO SE	قية الجديدة	
	$\theta_1 = \theta_1 - \infty$ (th) $\stackrel{\circ}{\underset{i=1}{\sum}} (h_{i} \otimes - y) (x_i)$	
	G, = 0 -0.0002 ((g) (	
	0,=0.02584375 قيار الحبية	
1=5		
	النطوات : الخطوات : الخطو	
1=3		
	0,= 0.0505921	
7-11	0,=0.0774678	
,	NY - WAY - THE TELEVISION	
1=2004		
	D. = 5.425425	
THE N	0,=76.406406	
	Y= Θ0 + Θ1 X : - 101- 4 coo	*
To late	Y=5.425 + (76.406)(4) = 98.1	
A STATE OF THE PARTY OF THE PAR		

#### Part 2:

```
| File | Edn | Selection | View | Go | Run | Terminal | Help | Go | Selection | Selection | Go | Go | Run | Terminal | Help | Go | Selection | Go | Go | Run | Terminal | Help | Go | Selection | Go | Go | Run | Terminal | Help | Go | Selection | Go | Go | Go | Run | Terminal | Help | Go | Selection | Go | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Go | Selection | Go | Run | Terminal | Help | Termi
```

