* سؤال 1:

الانحدار الخطي هو طريقة إحصائية تُستخدم لنمذجة العلاقة بين متغير تابع واحد أو أكثر وبين المتغيرات المستقلة. يفترض الانحدار الخطي وجود علاقة خطية بين المتغيرات المستقلة (الميزات) والمتغير التابع (الهدف).

الفكرة وراء الانحدار الخطي:

- التقريب بخط مستقيم: يحاول الانحدار الخطي إيجاد أفضل خط مستقيم يصف العلاقة بين المتغيرات المستقلة والتابعة. على سبيل المثال، في الانحدار الخطي البسيط مع متغير واحد، يتم البحث عن الخط الذي يتناسب أفضل مع نقاط البيانات على رسم بياني ثنائي الأبعاد.

- تقليل الخطأ: الهدف هو تقليل الانحرافات الرأسية (الفروق) بين القيم المتوقعة والبيانات الفعلية. يتم ذلك عادةً عن طريق تقليل دالة التكلفة، غالباً باستخدام طرق مثل أقل مربعات الفروق للعثور على الخط الذي يقلل من مجموع الفروق المربعية بين القيم المتوقعة والفعلية.

لماذا الانحدار بدلاً من التصنيف؟

- طبيعة المشكلة: يستخدم الانحدار عندما يكون المتغير التابع مستمرًا ورقميًا، حيث يحاول التنبؤ بقيم داخل نطاق معين (مثل توقع الأسعار أو درجات الحرارة). بينما يستخدم التصنيف عندما يكون المتغير التابع تصنيفيًا، يهدف إلى تصنيف المشاهدات في فئات محددة.

- تفسير النتائج: يوفر الانحدار إخراجًا كميًا، مما يجعله مناسبًا لسيناريوهات تحتاج إلى تنبؤات رقمية محددة. بينما يتعامل التصنيف مع نتائج تصنيفية، حيث يقوم بتصنيف الحالات في فئات مختلفة.

- أهداف تحليلية مختلفة: يخدم الانحدار والتصنيف أهداف تحليلية مختلفة ويستخدمان في مجالات مختلفة استنادًا إلى طبيعة المشكلة. على سبيل المثال، توقع أسعار العقارات (انحدار) مقارنة بتصنيف ما إذا كان البريد الإلكتروني هو رسالة مزعجة أم لا (تصنيف).

كلا الانحدار والتصنيف أساسيان في تعلم الآلة وتحليل البيانات، وتوجد قابلية تطبيقهما بناءً على طبيعة وأهداف المشكلة.

الفكرة الأساسية وراء الانحدار الخطي هي العثور على خط أو مستوى فاصل يتناسب بشكل أفضل مع البيانات، والذي يُقلل الفروقات بين القيم الفعلية والمتوقعة.

يُستخدم الانحدار عندما يكون الإخراج قيمة مستمرة (مثل التنبؤ بالأسعار أو درجات الحرارة)، بينما يُستخدم التصنيف عندما يكون الإخراج تصنيفيًا (مثل تحديد ما إذا كان بريد إلكتروني هو رسالة غير مرغوب فيها أو لا، أو التنبؤ بفئات مثل القطط أو الكلاب).

الانحدار الخطي يهدف إلى تقليل الفروق بين القيم المتوقعة والقيم الفعلية من خلال ضبط معاملات المعادلة الخطية. يستمر هذا العمل حتى يجد النموذج الخط الأفضل الذي يُمثل العلاقة بين المتغيرات.

الانحدار الخطي هو طريقة إحصائية تُستخدم لنمذجة العلاقة بين متغير يعتمد على عوامل أخرى وبين متغيرات مستقلة واحدة أو أكثر، وذلك من خلال تناسب معادلة خطية مع البيانات المرصودة. يفترض الانحدار الخطي وجود علاقة خطية بين المتغيرات، ويهدف إلى إيجاد الخط الأفضل الذي يُمثل هذه العلاقة ويُقلل من الاختلاف بين القيم المتوقعة والفعلية.

الفكرة الأساسية وراء الانحدار الخطي هي البحث عن خط مستقيم يمثل العلاقة بين المتغيرات المستقلة والمتغير المعتمد بشكل جيد. هذا الخط يُحدد من خلال حساب ميل الخط وقيمة التقاطع بحيث يكون الفرق بين البيانات الفعلية والخط الأقلّ مربعاً.

تخيل رسم نقاط البيانات على محورين، المحور الأفقي يمثل المتغيرات المستقلة والمحور العمودي يمثل المتغير المعتمد. الانحدار الخطي يسعى لرسم خط مستقيم يمر بين هذه النقاط ويمثل الاتجاه العام للعلاقة بين المتغيرات بحيث يكون الاختلاف بين الخط والبيانات الفعلية أقل قدر ممكن.

أما بالنسبة للسؤال عن سبب استخدام الانحدار بدلاً من التصنيف، فالاختيار يعتمد على طبيعة المشكلة. يُستخدم الانحدار عندما يكون المخرج أو المتغير المعتمد مستمرًا، مثل توقع أسعار العقارات أو درجات الحرارة. بينما يُستخدم التصنيف عندما يكون المخرج مجزأً إلى فئات مختلفة، حيث يهدف إلى تصنيف البيانات إلى فئات محددة. على سبيل المثال، توقع ما إذا كان البريد الإلكتروني هو رسالة سخام أم لا هو مشكلة تصنيفية، بينما توقع سعر منزل استنادًا إلى ميزاته هو مشكلة انحدارية. كل تقنية تناسب أنواع مختلفة من المهام التنبؤية استنادًا إلى طبيعة المتغير المعتمد.

سوال2

توضيح خوارزمية الانحدار الخطي: نظرة عامة على خوارزمية الانحدار الخطي:

- البداية: قم بتهيئة معلمات النموذج (ثيتا) بقيم عشوائية أو قيم صفرية. حدد معدل التعلم (ألفا) وعدد التكرارات أو معيار الانحسار للتوقف.

- التكرارات: قم بتكرار عملية التدريب لعدد محدد من التكرارات أو حتى الوصول إلى الانحسار.

- التنبؤ: حسب القيم المتوقعة عن طريق ضرب الميزات الداخلية (X) بالمعلمات (ثيتا).

- حساب التكلفة: قم بحساب دالة التكلفة، وعادةً باستخدام معيار متوسط مربعات الخطأ (MSE) أو قياسات مناسبة أخرى.

- حساب التدرج: قم بحساب التدرجات عن طريق إيجاد الاشتقاق لدالة التكلفة مع العوامل.

- تحديث المعلمات: قم بتحديث المعلمات (ثيتا) عن طريق طرح حاصل ضرب معدل التعلم والتدرجات.

- فحص الانحسار: تحقق مما إذا كانت معايير الانحسار تم تحقيقها (على سبيل المثال، تغير طفيف في التكلفة). إذا تم تحقيقها، قم بإيقاف التكرار.

- العودة: أعد المعلمات المتعلمة (ثيتا).

# توضيح حول الانحدار الخطي:

الانحدار الخطي هو أسلوب إحصائي أساسي يُستخدم لنمذجة العلاقة بين متغير أو أكثر مستقل ومتغير تابع. يفترض وجود علاقة خطية بين الميزات الداخلية والمخرجات.

تهدف الخوارزمية إلى إيجاد أفضل معادلة خطية تصف البيانات، وتقليل الفارق بين القيم المتوقعة والفعلية عن طريق تعديل المعلمات (ثيتا) بشكل تكراري باستخدام الانحدار التدرجي، متجهاً نحو الحد الأدنى لدالة التكلفة.

المعلمات المتعلمة (ثيتا) التي تم الحصول عليها من عملية التدريب يمكن استخدامها للتنبؤ بالمتغير التابع لبيانات جديدة.

الانحدار الخطي يُستخدم على نطاق واسع في مختلف المجالات لمهام مثل التنبؤ بالأسعار، وتقدير الاتجاهات، أو فهم العلاقات بين المتغيرات.

**2)**

**Pseudocode**

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:

// Computing predicted values

Predicted = theta0 + x \* theta1

// Calculating cost function with (MSE)

Calculate cost = (1 / 2 \* m) \* sum((Predicted - y)^2)

derivative\_ theta0 = (1 / n) \* sum(Predicted - y)

// Calculate derivative

derivative\_ theta1 = (1 / n) \* sum((Predicted - y) \* x)

// Update parameters

theta0 = theta0 - alpha \* derivative\_ theta0

theta1 = theta1 - alpha \* derivative\_ theta1

// Return updated parameters theta0, theta1

Return theta0, theta1

3)

لنفترض أنك تستخدم Batch Gradient Descent وقمت برسم خطأ التحقق من الصحة في كل فترة. إذا لاحظت أن خطأ التحقق من الصحة يتزايد باستمرار، فما الذي يحدث على الأرجح؟ كيف يمكنك اصلاح هذا؟

إذا كان خطأ التحقق من الصحة يتزايد باستمرار في كل فترة عند استخدام Batch Gradient Descent، فهذا يشير إلى حالة التجهيز الزائد. يحدث التجاوز عندما لا يتعلم النموذج الأنماط الأساسية فحسب، بل يتعلم أيضًا الضوضاء والخصائص المحددة لبيانات التدريب، مما يؤدي إلى أداء جيد في مجموعة التدريب ولكن بشكل سيئ على البيانات الجديدة غير المرئية (مثل مجموعة التحقق من الصحة).

# الأسباب المحتملة:

1. نموذج معقد: قد يكون النموذج معقدًا جدًا، حيث يلتقط تشويشًا أو تفاصيل غير ضرورية من بيانات التدريب التي لا يتم تعميمها بشكل جيد على البيانات الجديدة.

2. معدل التعلم الكبير: يمكن أن يؤدي معدل التعلم المرتفع إلى تجاوز الحد الأدنى لدالة التكلفة، مما يجعل من الصعب على النموذج أن يتقارب مع الحل الأمثل.

# الإصلاحات المحتملة:

1. تقنيات التنظيم: تقديم أساليب التنظيم مثل تنظيم L1 أو L2 لمعاقبة النماذج المعقدة للغاية. تضيف هذه التقنيات عقوبة جزائية إلى دالة التكلفة، مما يمنع النموذج من أن يصبح معقدًا للغاية.

2. تقليل تعقيد النموذج: تبسيط النموذج عن طريق تقليل عدد الميزات، أو إجراء اختيار الميزات، أو تقليل سعة النموذج (على سبيل المثال، تقليل عدد الطبقات أو العقد في الشبكة العصبية).

3. ضبط معدل التعلم: قم بتقليل معدل التعلم للسماح لعملية التحسين بالتقدم بشكل تدريجي، مما قد يؤدي إلى منع التجاوز والسماح للنموذج بالتقارب بشكل أكثر سلاسة.

4. التوقف المبكر: قم بتنفيذ التوقف المبكر من خلال مراقبة خطأ التحقق من الصحة أثناء التدريب. قم بإيقاف التدريب عندما يبدأ خطأ التحقق من الصحة في التزايد باستمرار، مما يمنع النموذج من الإفراط في التناسب مع بيانات التدريب.

5. التحقق المتبادل: استخدم تقنيات مثل التحقق المتبادل لتقييم أداء تعميم النموذج بشكل أفضل وضبط المعلمات الفائقة وفقًا لذلك.

من خلال تطبيق التنظيم، وضبط تعقيد النموذج، وضبط المعلمات الفائقة، ومراقبة خطأ التحقق من الصحة عن كثب، يمكنك تخفيف التجاوز وتحسين قدرة النموذج على التعميم بشكل جيد على البيانات غير المرئية.

إذا زادت أخطاء التقييم الخاصة بمجموعة الاختبار تدريجياً في كل دورة (epoch) خلال استخدام خوارزمية Batch Gradient Descent، فهذا يشير إلى أن النموذج يعاني من مشكلة الـ"Overfitting" أو التعميم الزائد. الـOverfitting يحدث عندما يتعلم النموذج البيانات التدريبية بشكل زائد، حيث يتعلم تفاصيل صغيرة وضوضاء قد تكون غير مفيدة للتنبؤ بالبيانات الجديدة أو مجموعة الاختبار.

أسباب زيادة أخطاء التقييم:

١. \*\*نموذج معقد\*\*: قد يكون النموذج معقداً جداً بالنسبة لحجم البيانات، مما يسمح له بالتعلم من الضوضاء بدلاً من الأنماط الحقيقية.

٢. \*\*نقص التنظيم\*\*: ربما لا يتم تطبيق تقنيات التنظيم الجيدة مثل L1/L2 regularization أو dropout، مما يؤدي إلى انحراف النموذج.

٣. \*\*حجم البيانات المحدود\*\*: قد يكون حجم البيانات صغيراً، الأمر الذي يجعل النموذج يتعلم تفاصيل محددة بدلاً من العامة.

طرق التعامل مع زيادة أخطاء التقييم:

١. \*\*التنظيم\*\*: زيادة التنظيم عن طريق إضافة عوامل تقنينية مثل L1 أو L2 regularization للنموذج لمنع التعميم الزائد.

٢. \*\*تقليل تعقيد النموذج\*\*: تقليل تعقيد النموذج عن طريق تقليل عدد المعلمات أو استخدام نماذج بسيطة لتمثيل الأنماط البيانية.

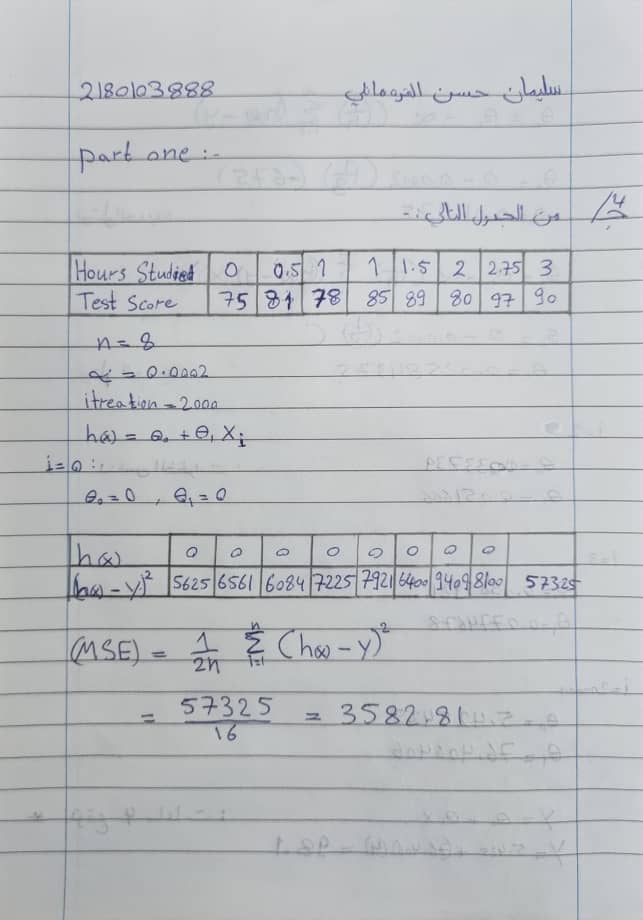
٣. \*\*زيادة البيانات\*\*: زيادة حجم البيانات من خلال تقنيات إثراء البيانات لتوفير مثالات متنوعة.

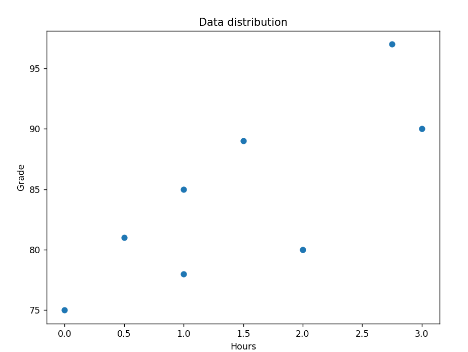
٤. \*\*إيقاف مبكر\*\*: تطبيق إيقاف مبكر، حيث يتم إيقاف التدريب إذا بدأت أخطاء التقييم بالارتفاع بشكل متكرر، وذلك لمنع النموذج من التعميم الزائد.

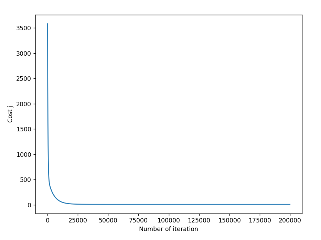
٥. \*\*التقييم المتقاطع\*\*: استخدام تقنيات مثل التقييم المتقاطع لتقييم أداء النموذج على مجموعات بيانات مختلفة والتأكد من عدم استمرار مشكلة التعميم الزائد.

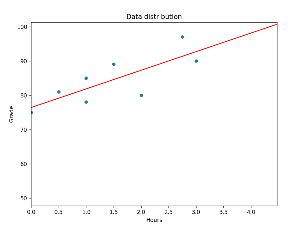
من خلال التعامل مع هذه القضايا، يمكنك تحسين قدرة النموذج على التعميم وتقليل الـOverfitting، وبالتالي تحسين أداءه على مجموعة الاختبار أو التقييم.

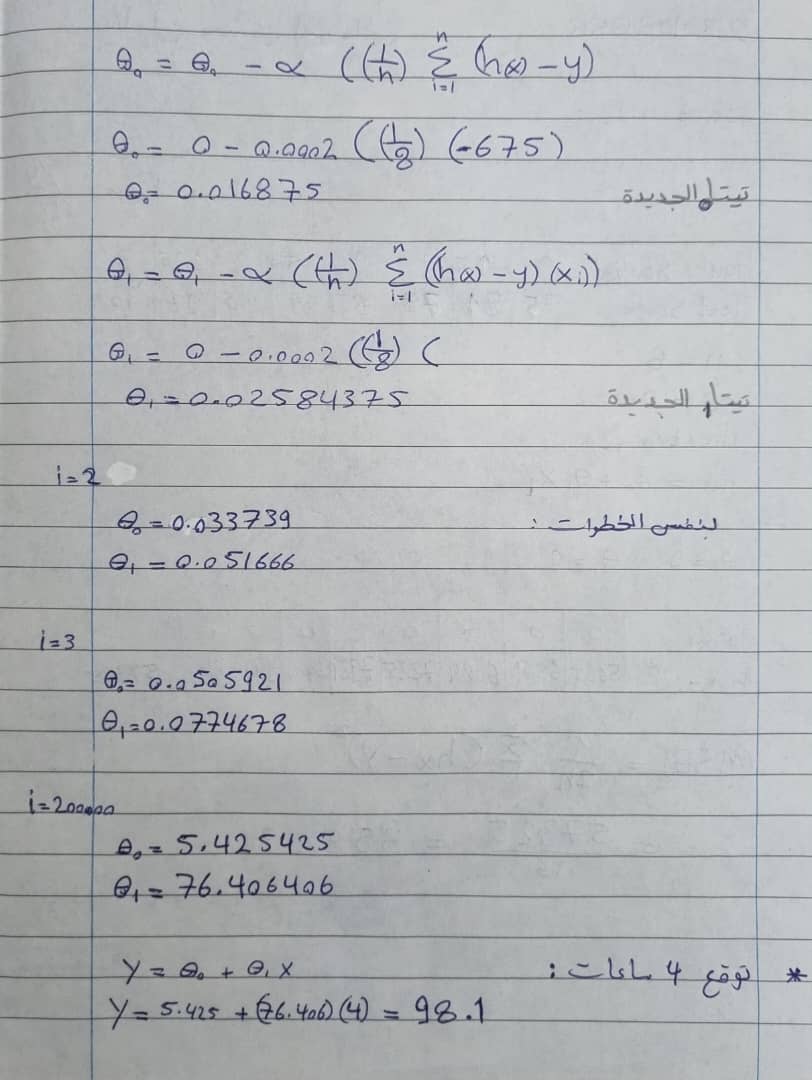
4)











Part 2:

