

نماذج الانحدار في التعلم الآلي: شرح موجّه لطلاب اللسانيات

إعداد د. محمد بن صويلح الزايدى

<https://faculty.ksu.edu.sa/ar/malzaidi1/cv>

النماذج الانحدارية هي أدوات تُستخدم في التعلم الآلي (Machine Learning) لتحليل العلاقات بين المتغيرات والتنبؤ بقيم مستمرة (أرقام) بناءً على بيانات سابقة. في مجال اللسانيات، يمكن استخدامها لتحليل الأنماط الصوتية (مثل التردد الأساسي F0 أو مدة النطق (Duration)) وفهم العلاقات بين الخصائص اللغوية (مثل ال Stress/Intensity أو الموقع النحوي). أمثلة توضيحية في اللسانيات:

1. التنبؤ بقيمة التردد الصوتي بناءً على موضع الكلمة في الجملة.
2. التنبؤ بالمدة الزمنية لنطق كلمة بناءً على عدد المقاطع ومدى ال Stress
3. تحليل الفروق في النغمة الصوتية بين الجمل الخبرية والجمل الاستفهامية.

أنواع النماذج الانحدارية

لنلق نظرة على الأنواع الرئيسية للنماذج الانحدارية مع أمثلة من اللسانيات:

• الانحدار الخطي (Linear Regression)

يحاول النموذج رسم خط مستقيم يوضح العلاقة بين متغيرين.

مثال: التنبؤ بمدة نطق الكلمة بناءً على موقعها في الجملة. فمثلاً: الكلمات في بداية الجملة قد تكون أقصر مدة مقارنة بالكلمات في نهايتها.

يحسب النموذج أفضل خط مستقيم يقلل الفارق بين القيم الفعلية (مدة النطق) والقيم المتنبأ بها.

• الانحدار المتعدد (Multiple Regression)

يوسع نموذج الانحدار الخطي ليشمل عوامل متعددة.

مثال: التنبؤ بالتردد الصوتي (F0) بناءً على نوع الجملة (خبرية أو استفهامية)، وعدد المقاطع في الكلمة، وموقعها في الجملة.

النموذج يحلّل مدى تأثير كل عامل على التردد الصوتي النهائي.

• الانحدار متعدد الحدود/كثير الحدود (Polynomial Regression)

يضيف منحنيات للعلاقة بين المتغيرات.

مثال: التنبؤ بمنحنى النغمة الصوتية (pitch contour) في الجملة الاستفهامية.

العلاقة بين التردد الصوتي والزمن قد تكون منحنية، حيث يبدأ التردد بالارتفاع تدريجيًا ثم ينخفض بنمط منحنى.

يمكن استخدام هذا النوع لتحليل التغيرات الصوتية الديناميكية في الحوارات أو النصوص الشعرية.

• انحدار Ridge وLasso

كلاهما يُستخدم لمنع الإفراط في التكيّف والمعروف باسم (Overfitting)، وهو ما يحدث عندما يصبح النموذج معقدًا جدًا لدرجة أنه يؤدي بشكل جيد على البيانات التدريبية لكنه يفشل مع البيانات الجديدة.

- انحدار Ridge يقلل من حجم القيم العددية (المعاملات) لكنه لا يحذفها.
- مثال: التنبؤ بطول الحركات الصوتية مع ضمان ألا يعتمد النموذج بشكل مفرط على عامل واحد فقط (مثل ال stress فقط دون أخذ الموقع في الجملة في الاعتبار).
- انحدار Lasso يقلل بعض القيم العددية إلى الصفر تمامًا، مما يعني حذف العوامل غير المهمة.
- مثال: إذا تبين أن "سرعة المتحدث" ليست ذات تأثير كبير على التنبؤ بالتردد الصوتي، يمكن للنموذج تجاهلها.

• الانحدار المرن (Elastic Net Regression)

يجمع بين ميزات Ridge وLasso.

مثال: إذا أردنا أن يأخذ النموذج في الاعتبار جميع العوامل (مثل ال stress، الموقع، وعدد المقاطع) مع التركيز فقط على الأكثر أهمية.

• الانحدار غير الخطي (Nonlinear Regression)

يتعامل مع العلاقات الأكثر تعقيدًا التي لا يمكن تمثيلها بخطوط مستقيمة أو منحنيات بسيطة.

مثال: تحليل العلاقة بين درجة الصوت (pitch) وسرعة الكلام في سياقات مختلفة حيث قد تتغير العلاقة بطرق غير متوقعة وغير خطية.

يمكن استخدام هذا النوع في دراسة تأثير السياق النحوي أو السياق الاجتماعي على النغمة الصوتية.

التحديات والقيود في استخدام النماذج الانحدارية في اللسانيات

1. حجم البيانات: النماذج الانحدارية تتطلب بيانات كافية للتعلم. في الدراسات اللسانية، جمع البيانات الصوتية بدقة قد يكون تحديًا.

2. الإفراط في التكيّف: مع وجود عدد كبير من العوامل، قد يصبح النموذج مفرط التخصص في بيانات التدريب ويؤدي بشكل ضعيف مع البيانات الجديدة.
3. التعقيد الحسابي: النماذج غير الخطية والمرنة تحتاج إلى تقنيات حسابية متقدمة.