



# معسكر علم البيانات و تعلم الآلة

2022 - 11 - 21



## نبذة عن المدرب



# محتوى المعسكر

اليوم	الأسبوع الأول Getting Started	الأسبوع الثاني Data Analysis and Visualization	الأسبوع الثالث Machine Learning	الأسبوع الرابع EDA & FE in Action	الأسبوع الخامس Modeling Interpretation in Action	الأسبوع السادس Final Project
الأحد	Intro to DS	NumPy	Intro to ML	DS Knowledge Catalog	Models Families: Distance & Time Series	Final Project
الاثنين	Git & Github	Pandas	Supervised ML	EDA1: Univariate & Multivariate Analysis	Models Evaluation: Regression & Classification	Final Project
الثلاثاء	Python Review	Matplotlib	Supervised ML	EDA2: Association Analysis & Hypothesis Construction	Optimization Techniques	Final Project
الأربعاء	Python Review	Seaborn	Unsupervised ML	Features Engineering: Scaling, Merging & Discretization	NLP and Text Mining Basics	Final Project
الخميس	Python Review	Plotly	Unsupervised ML	Models Families: Continuous & Categorical	Neural Networks Basics	Presentation

**\*\*ملاحظة: قد تتغير المواضيع أو أوقات طرحها بناء على تقدم الطلاب.**



# تقييم النماذج



## تقييم النماذج

مرحلة نهدف فيها قياس فاعلية، قوة، وضعف النماذج التي تم بناؤها  
في مرحلة التطوير

تساعدنا في اتخاذ القرار باعتماد نسخة النموذج أو إعادة تطويره/تحسينه



# أنواع تقييم النماذج

## تقييم نماذج الانحدار Regression

- \* Relative Error
- \* Absolute Error
- \* Correlation



## تقييم نماذج التصنيف Classification

- \* Precision
- \* Recall
- \* Accuracy





# تقييم نماذج الانحدار

## Regression



# Absolute Error

المعادلة:

$$|y - Y|$$

- هو قياس الفرق **المطلق** بين القيم الفعلية إلى القيم المتوقعة

مثلاً: توقع سعر منتج س = 400 والقيمة الفعلية للمنتج = 380،

فما هو الخطأ المطلق هنا؟

$$20 = |400 - 380|$$





# Relative Error

المعادلة:

$$\frac{|y - Y|}{Y}$$

- هو قياس الفرق النسبي بين القيم الفعلية إلى القيم المتوقعة

مثلاً: توقع سعر منتج س = 400 والقيمة الفعلية للمنتج = 380،

فما هو الخطأ المطلق هنا؟

$$0.053 = 280 / 20 = |400 - 380| / 380$$

$$\text{والنسبة المئوية } 5.3\% = 100 * 0.053$$



# Mean Absolute Error - MAE

المعادلة:

$$\frac{1}{n} \sum_{i=1}^n |y_i - Y_i|$$

- هو قياس الفرق **النسبي** بين القيم الفعلية إلى القيم المتوقعة
- كلما قلت قيمة MAE كلما كان النموذج أفضل

مثلاً: مجموع الخطأ المطلق لتوقعات أسعار المنتجات = 230

وعدد التوقعات = 10، فما هي قيمة MAE هنا؟

$$23.0 = 230/10$$

# Root Mean Squared Error - RMSE

المعادلة:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - Y_i)^2}$$

- هو قياس متوسط الخطأ لتنبؤات النموذج
- كلما قلت قيمة RMSE كلما كان النموذج أفضل

مثلاً: مجموع الخطأ التربيعي لتوقعات أسعار المنتجات = 8100

وعدد التوقعات = 10، فما هي قيمة RMSE هنا؟

$$28.46 = \text{sqrt}(810) = \text{sqrt}(8100/10)$$



# تقييم نماذج التصنيف

## Classification



# Precision

المعادلة:

$$\frac{TP}{TP + FP}$$

- قيمة تساعدنا بمعرفة نسبة عدد الحالات التي تم التنبؤ بها بشكل صحيح وتبين أنها إيجابية فعلاً

- تُعدّ مقياس مفيد في الحالات التي يكون فيها FP مثير للاهتمام أكثر من FN

# Recall

المعادلة:

$$\frac{TP}{TP + FN}$$

- قيمة تساعدنا بمعرفة نسبة عدد الحالات الإيجابية الفعلية التي تمكن النموذج من التنبؤ بها بشكل صحيح
- يُعد مقياسًا مفيدًا في الحالات التي تتطلب اهتمامًا أكبر بالحالات FN عن FP

حالات يُرَجَّح فيها اعتماد Recall عن Precision ؟



# Accuracy

المعادلة:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- هي دقة التوقعات مقارنة بالقيم الفعلية



# F1 Score

المعادلة:

$$\frac{2TP}{2TP + FP + FN} \quad \text{OR} \quad 2 * \frac{Precision \cdot Recall}{Precision + Recall}$$

- هي قيمة تجمع بين أهمية القيمتين Precision والRecall
- تكون بقيمتها الأمثل إذا كان  $Recall = Precision$
- لتفسيرها بالطريقة الأفضل لمعرفة الأكثر تأثيرًا يُنصح بحساب Precision والRecall أولًا لمعرفة القيمة المؤثرة على نتيجة الدقة

# AUC

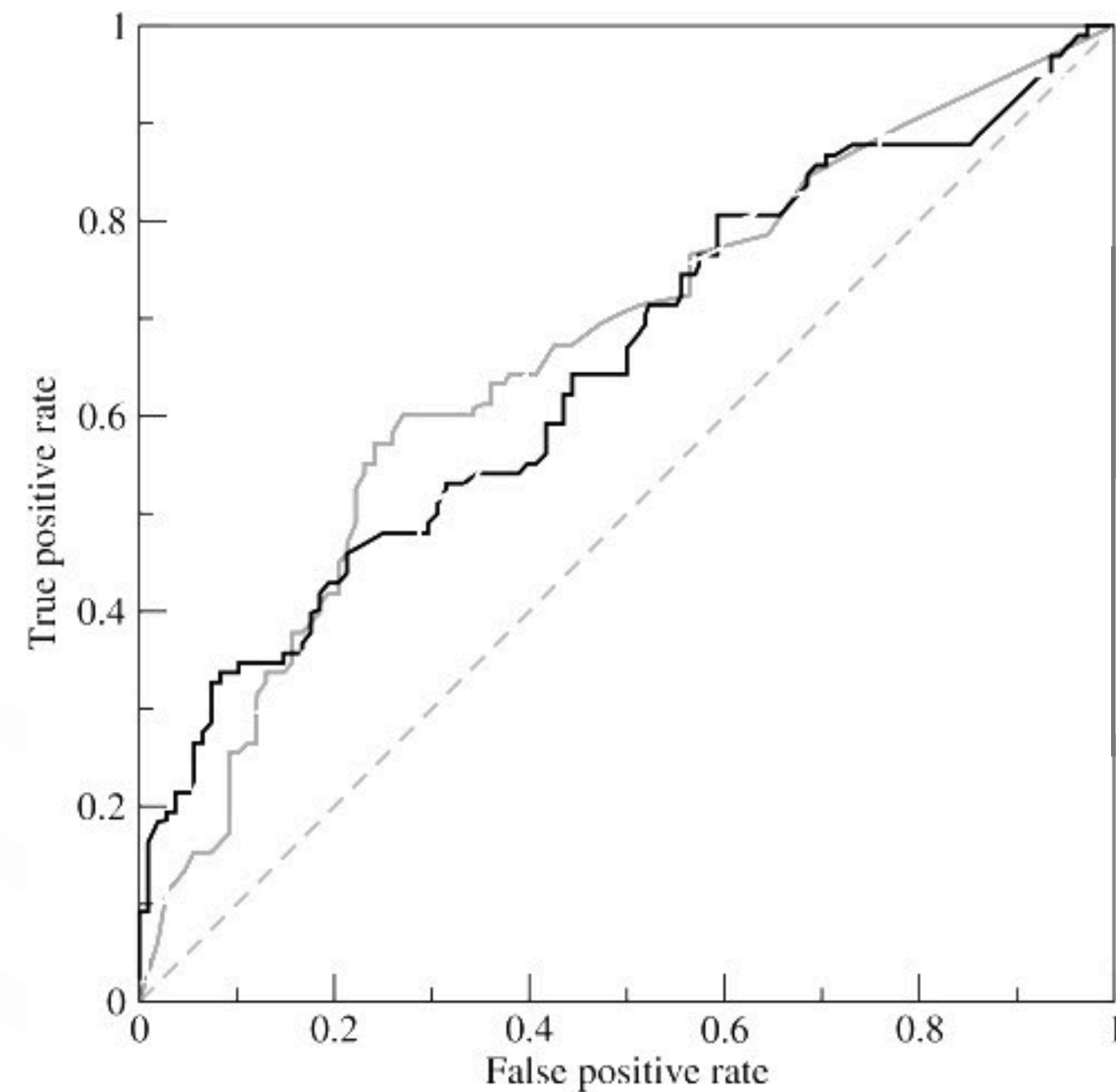
## المعادلات:

$$TPR = Sensitivity = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

OR

$$Specificity = \frac{TN}{TN + FP} \gg FPR = 1 - Specificity$$



المصدر: [https://en.wikipedia.org/wiki/Receiver\\_operating\\_characteristic](https://en.wikipedia.org/wiki/Receiver_operating_characteristic)



## للتسليم

في الجدول المقابل يظهر لنا نوعين من التوقع، تصنيف ورقم متوقع، والمطلوب التالي:

1. وضع قيم أخرى آخر خيالي كأداء نموذج آخر

3. حساب جميع معاملات الخطأ للنموذجين في للقيم في الجدول المقابل مرة، والقيم الجديدة مرة أخرى

5. المقارنة بين أداء النموذجين بالنظر إلى النسب المحسوبة

Hours Studied	Exam Score	Predicted Score - M1	Predicted Score - M2	Actual Pass?	Predicted Pass? - M3	Predicted Pass? - M4
1	68	79.03	?	N	Y	?
1	78	79.03	?	N	Y	?
1	75	79.03	?	N	Y	?
2	83	82.11	?	Y	Y	?
2	80	82.11	?	Y	Y	?
2	78	82.11	?	N	Y	?
2	89	82.11	?	Y	Y	?
2	93	82.11	?	Y	Y	?
3	90	85.19	?	Y	Y	?
3	91	85.19	?	Y	Y	?
4	94	88.27	?	Y	Y	?
5	88	91.35	?	Y	Y	?
5	84	91.35	?	Y	Y	?
5	90	91.35	?	Y	Y	?
6	94	94.43	?	Y	Y	?



استفساراتكم؟ 🤔