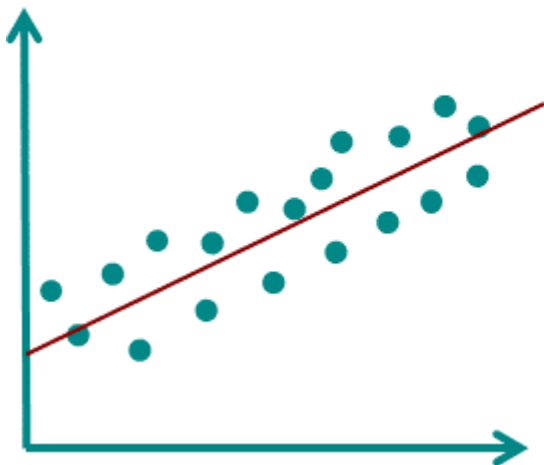


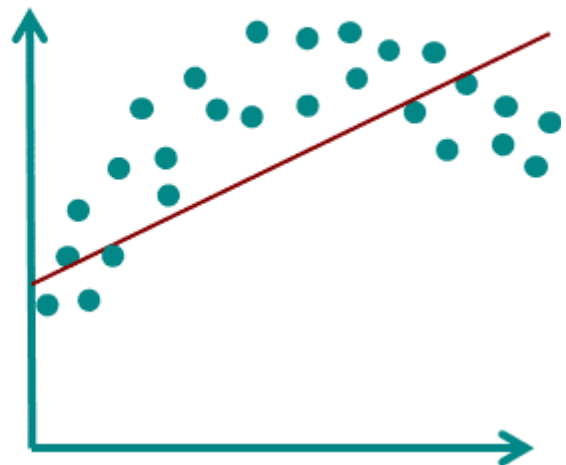
رگرسیون

- + دو نوع است = خطی و غیر خطی
- + برای مقادیر پیوسته استفاده میشه
- + معیار ارزیابی میزان خطا است
- + خطا = فاصله هر نمونه از رگرسیون میشه خطای ما

Linear



Non Linear



+ روش هایی که برای بدست آوردن میزان خطا داریم :

1. MAE (mean absolute error)

میانگین قدر مطلق ارور ها

$$MAE = (1/n) * \sum |y_{true} - y_{predict}|$$

- n : تعداد نمونه ها (داده ها)
- y_{true} : مقدار واقعی برای نمونه
- $y_{predict}$: مقدار پیش بینی شده توسط مدل برای نمونه

2. MSE (mean squared error)

در این روش ما خطای هر نمونه با توجه به فاصله ای که از رگرسیون داره ارزیابی میکنیم وقتی خطا ها را به توان دو میرسانیم در واقع برای خطا ها بزرگتر جریمه ی بیشتری متحمل میشویم

$$MSE = (1/n) * \sum (y_{true} - y_{predict})^2$$

- n : تعداد نمونه ها
- y_{true} : مقدار واقعی
- $y_{predict}$: مقدار پیش بینی شده توسط مدل

3. RMSE (root mean squared error)

در این روش برای تغییر نکردن واحد های اصلی که داریم از فرمول قبلی جزر میگیریم

$$RMSE = \sqrt{(1/n) * \sum (y_i - \hat{y}_i)^2}$$

معیار ارزیابی رگرسیون

R2_SCORE

- + برای میزان ارزیابی مدل رگرسیون از این استفاده می کنیم
- + در واقع یک معیاری هست که نشون میده مدل ما چقدر خوب کار میکنه
- + چیزی که این مقدار رابطه به ما میگه این هستش که ما یک سری مقدار مستقل داریم که میخواهیم باهاش متغیر وابسته رو شناسایی کنیم باهاش میزان ارزیابی مدل ساخته شده توسط اون مقادیر رو میریم و پیشبینی میکنیم
- + تغییرات متغیر وابسته هدف بر اساس متغیرهای مستقل ویژگی ها توضیح میده

$$R^2 = 1 - (\sum(y_{\text{true}} - y_{\text{pred}})^2) / (\sum(y_{\text{true}} - \text{mean}(y_{\text{true}}))^2)$$

و نتیجه ای که برای R2_SCORE میشه گرفت :

خروجی همیشه مقداری بین صفر تا یک هست

مقدار بین ۰ و ۱ •

- مدل تمام تغییرات داده را توضیح می ده: (بهترین مقدار) ۱
- مدل هیچ بهبودی نسبت به پیش بینی میانگین ندارد: ۰
- مقادیر منفی: مدل از یک خط افقی (میانگین) بدتر عمل می کند (معمولاً نشان دهنده مشکل در مدل است).

در ادامه میخواهیم راجب انواع رگرسیون ها صحبت کنیم
ما دو نوع رگرسیون داریم:

- + رگرسیون خطی
- + رگرسیون غیر خطی

و در آینده راجب هر دوی این ها صحبت خواهیم کرد