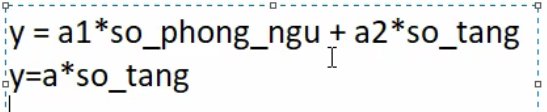
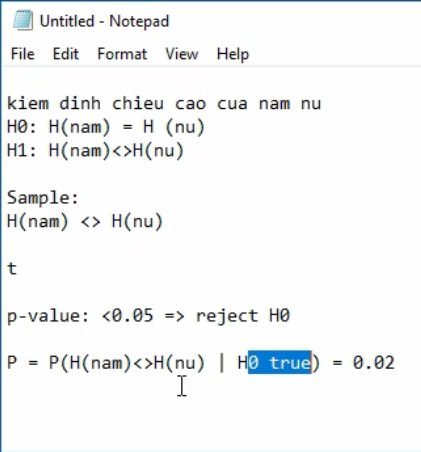


Robust regression: regression for outlier

Heteroscekasticity: residual include explanatory attributes

Multicolinear: when two attribute strongly relates





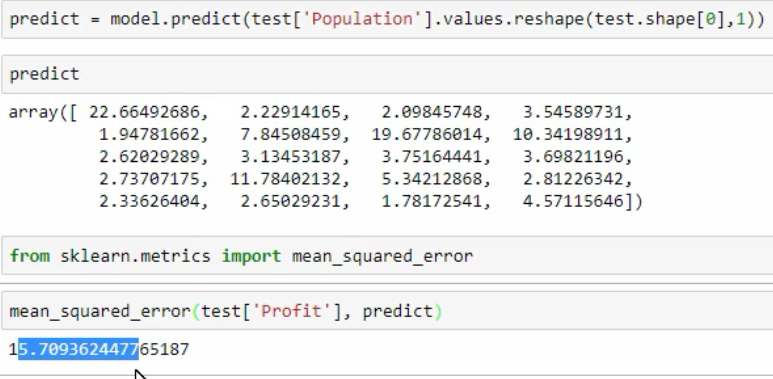
Given sample chiều cao nam khác chiều cao nữ. Câu hỏi: Lam sao biết trong population Chiều cao nam khác hay bằng chiều cao nữ

Đối với p value, cho H0 true (lấy ra 1 sample H0 true) proba mà H(nam) khác H(nu) la 0.02

* làm sao mà H0 true mà vẫn tìm ra H(nam) khác H(nu) -> H0 có thể không true

**Data nhiều chiều -> covariance matrix**

**Holdout:** kỹ thuật chia train test



Avoid overfitting: k-fold cross validation

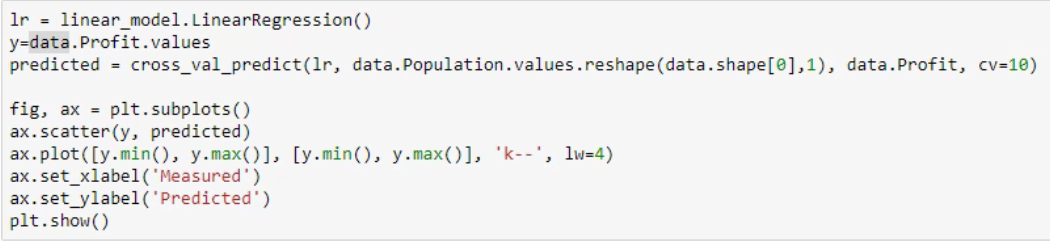


4 phần train 1 phần test -> model LG-> MSE

4 phần train khác, 1 phần test khác -> model LG-> MSE

…

Lấy model gần với MSE trung bình 5 lần

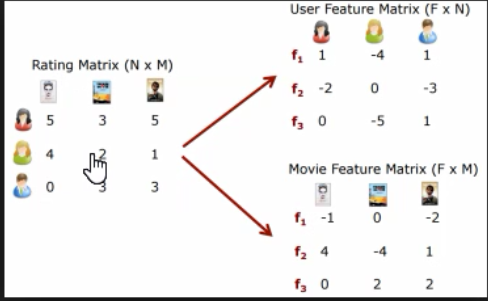


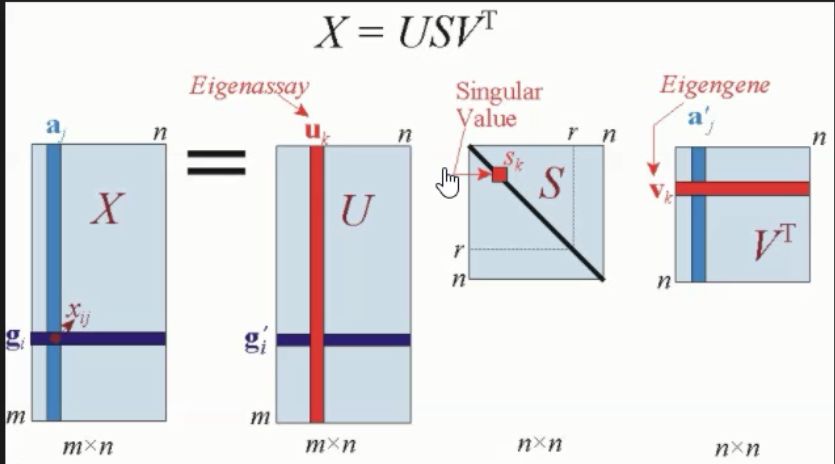
**Tuning Parameter Cross Validation**

**Avoid Multicolinear:**

PCA

**Partial least square regression**: singular value decomposition (matrix factorization)





Vi X rat lon nen can decompose

Trong recommendation, S la latent feature (feature ẩn)

Hoc regression tren S better tren X