

Agenda

- Feature Scaling impact on Loss Geometry
- Adjusted r^2 -score
- Sklearn vs statsmodel
- Assumptions of Linear Regression
 - Linear Relationship ✓
 - No multi-collinearity ✓
 - Normal Distribution of Residuals
 - Homoscedasticity
 - No auto-correlation



vas

@vasumanmoza

Follow



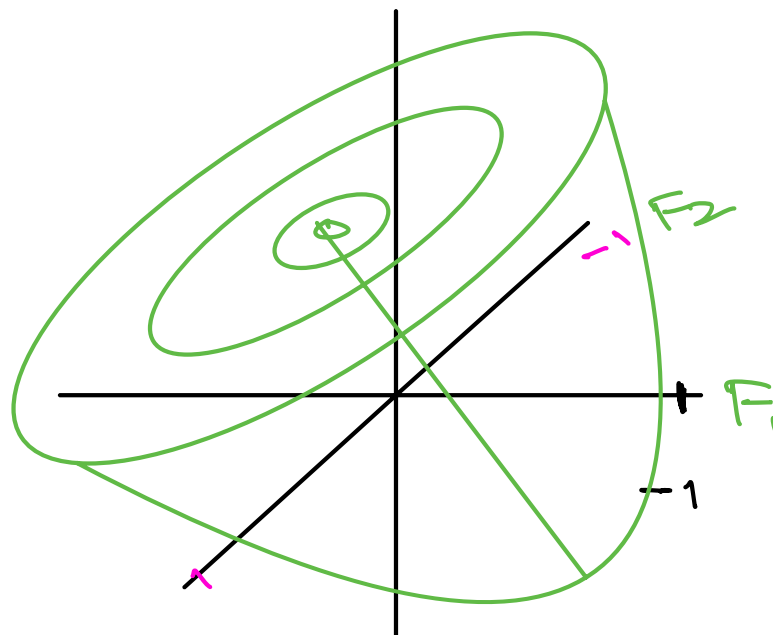
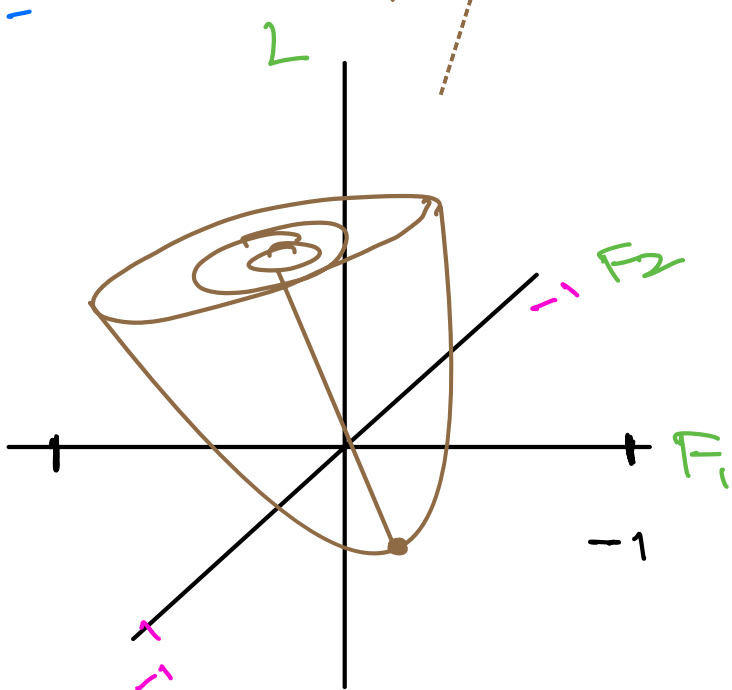
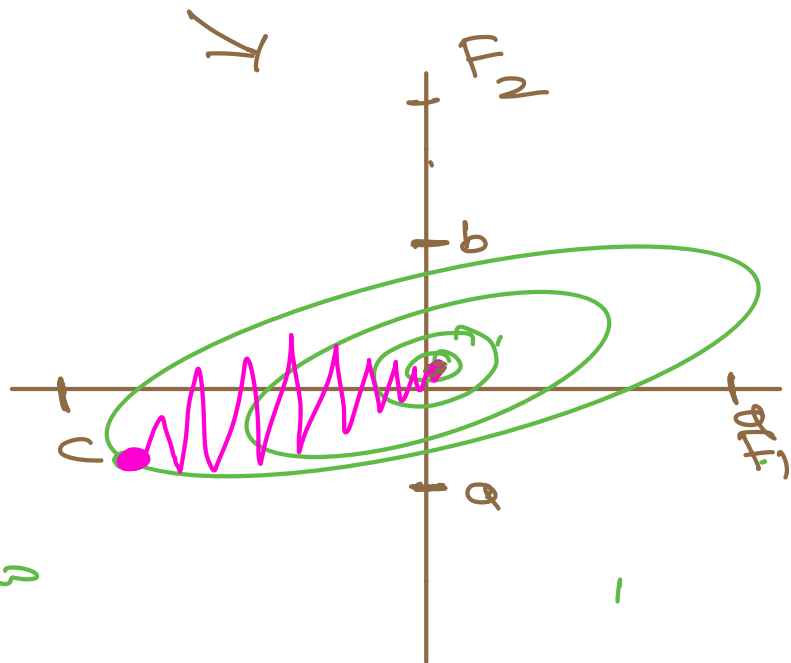
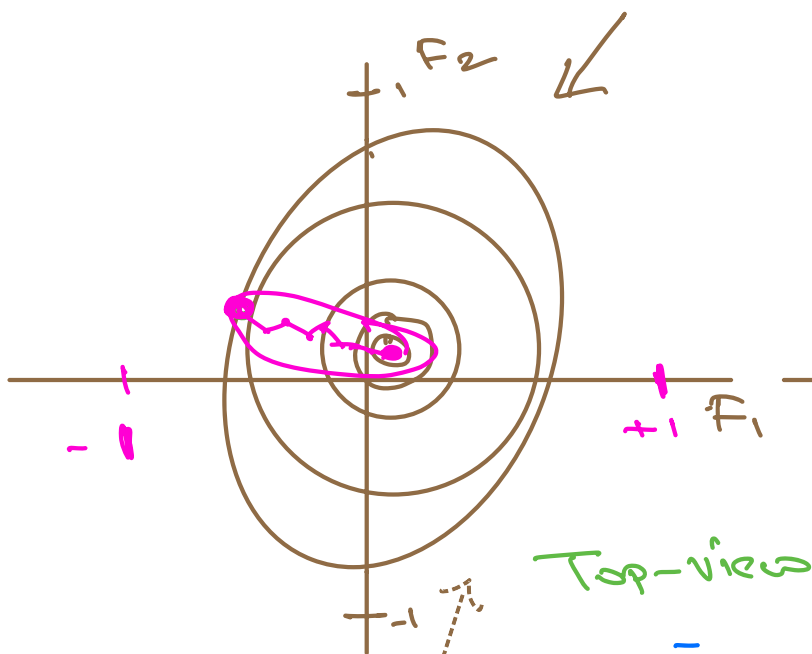
Claude 4 just refactored my entire codebase in one call.

25 tool invocations. 3,000+ new lines. 12 brand new files.

It modularized everything. Broke up monoliths. Cleaned up spaghetti.

None of it worked.
But boy was it beautiful.

Feature Scaling impact on Loss Fx Geometry



3D View of Loss

Adjusted r^2 -score

Case 1	Case 2
r^2 -score = 0.30	r^2 -score = 0.30
num-features = 2	num-features = 10

Occam's Razor

- When you have comparable performance, always pick the simple version

Note: r^2 -score either increase or stays same when features are increased

'if we add irrelevant feature

r^2 -score stays same

Adjusted R²-Score

$$1 - \left[\frac{(1 - R^2)(n - 1)}{(n - d - 1)} \right]$$

Annotations:
 - A green arrow points to R^2 .
 - A pink arrow points to $n - 1$ with the label "num-rows".
 - A pink circle is drawn around d in the denominator, with a pink arrow pointing to it from the label "num-feature".

$$(-\infty, 1)$$

(0, 1)

$$d + 1$$

If this dimension does not add significant value in Numerator

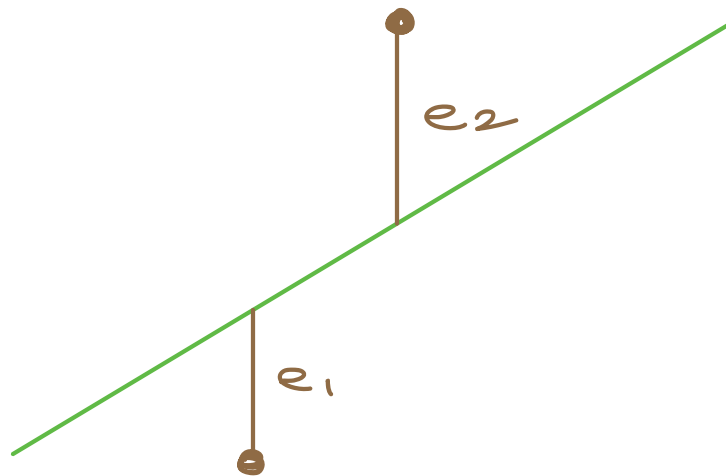
Adjusted R²-Score will Reduce

'if we add 'irrelevant' feature

adj R²-score will decrease

(H.W): Implement adj R²-Score in our Linear Regression Class

Stats model



Ordinary Least Squares

$$\min \left(\sum y \text{ Error}^2 \right)$$

Stats Model

- CI for features
- F-statistic (Model)
- P-Value (individual feature)
- residual Plot
- QQ Plot

Sklearn

- Less concerned with statistical detail
- Feature-scaling
- Regularization
- Cross-Validation
- Eval-metric

Assumptions of Linear Regression

- ① Linear Relationship
- ② No multi-collinearity
- ③ Normal Distribution of Residuals
- ④ Homoscedasticity
- ⑤ No auto-correlation

Por - Read