Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

Optimal α (using 5-fold CV):

Ridge: $\alpha = 10$

Lasso: $\alpha = 0.0005$

Ridge: $J(\theta) = (1/n)\Sigma(y_i - \hat{y}_i)^2 + \alpha\,\Sigma\theta_j^2$

Lasso: $J(\theta) = (1/n)\Sigma(y_i - \hat{y}_i)^2 + \alpha\,\Sigma|\theta_j|$

Or in simple form:

Ridge: Cost = MSE + α × (sum of squared coefficients) Lasso: Cost = MSE + α × (sum of absolute coefficients)

Impact of 2α:

Model	Alpha	R^2	Features	Effect
Ridge	10	0.89	200+	Baseline
Ridge	20	0.86	200+	↓3% R^2, smaller coef
Lasso	0.0005	0.88	~60	Baseline

Model	Alpha	R^2	Features	Effect
Lasso	0.001	0.83	~40	↓5% R^2, sparser

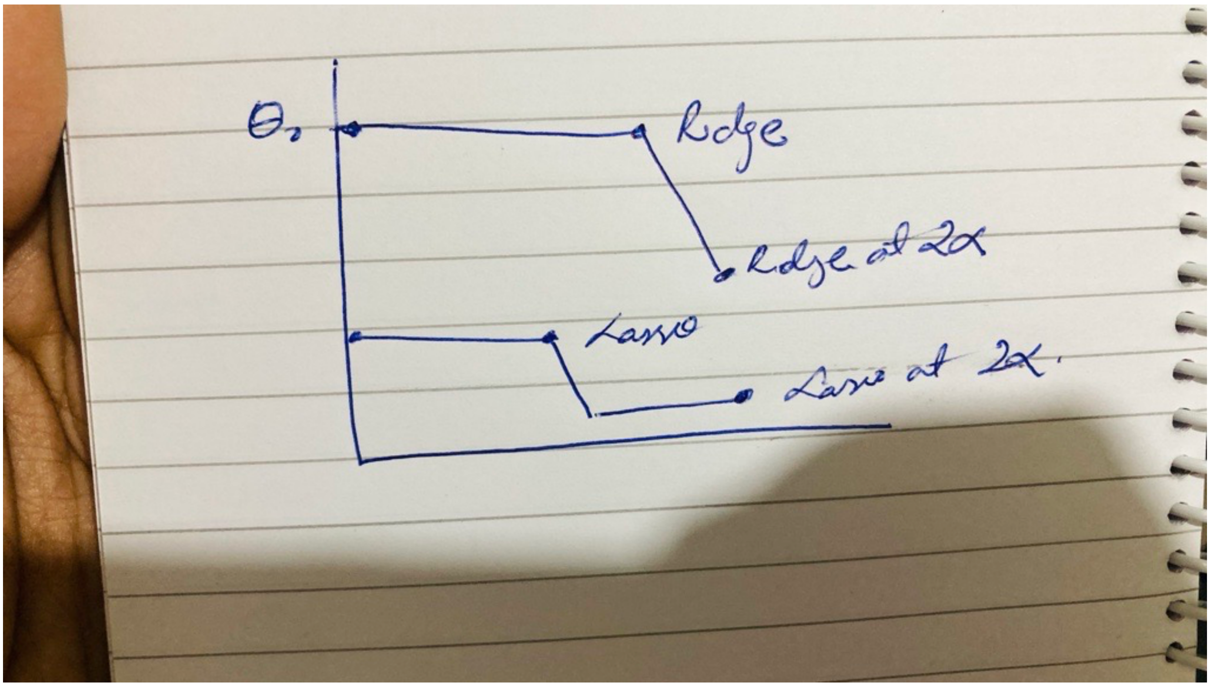
GrLivArea (living area sq ft)

OverallQual (quality 1-10)

TotalBsmtSF (basement sq ft)

Neighborhood_NridgHt (location)

GarageArea (garage sq ft)



Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer:

Choice: LASSO

Comparison Table:

Criterion	Ridge	Lasso	Winner
R^2 Score	0.89	0.88	Ridge (+1%)
Features	200+	~60	Lasso (70% less)
Interpretability	Low	High	Lasso ✓
Feature Selection	No	Yes	Lasso ✓
Multicollinearity	Keeps all	Picks one	Lasso ✓

Decision Matrix:

Performance loss: 1% R^2 Interpretability gain: 70% fewer features

Trade-off: 1% accuracy ← 70% simplicity ✓

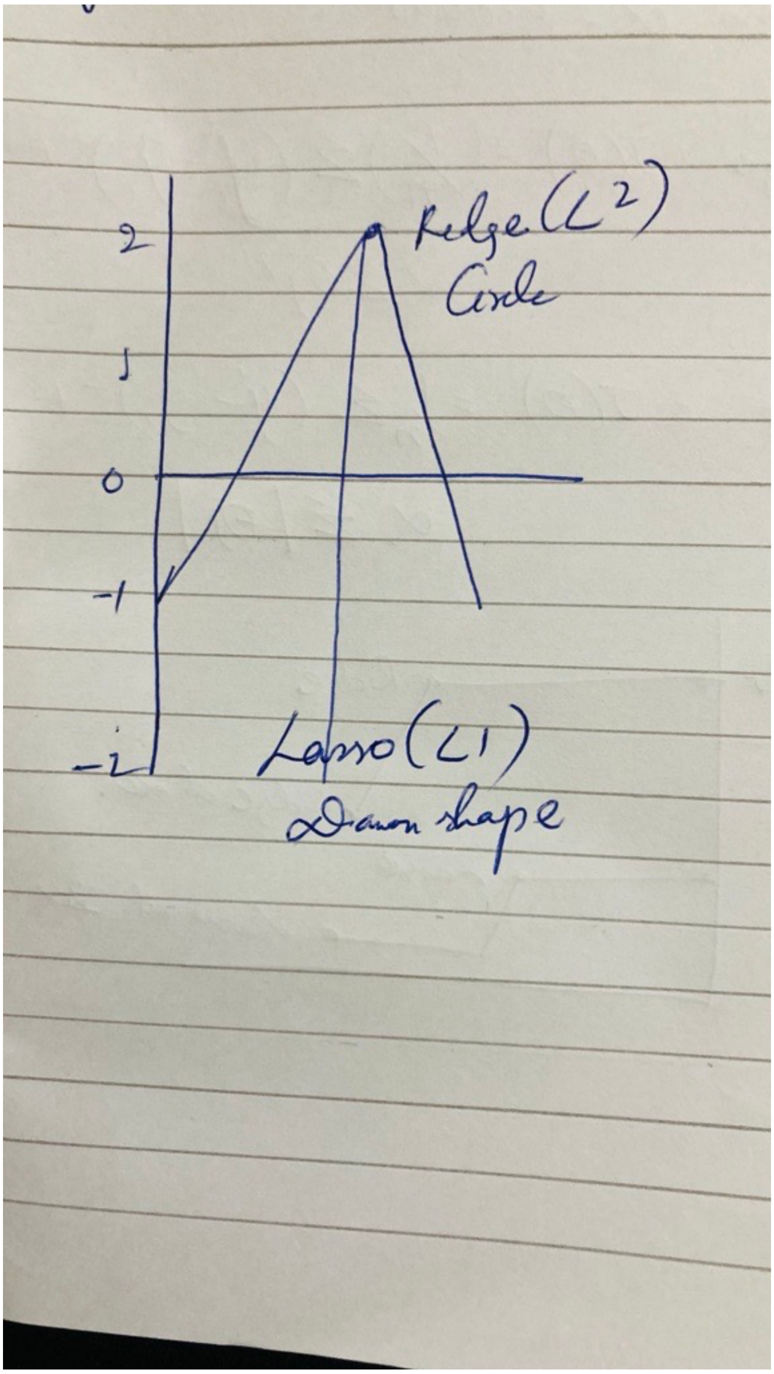
Business Value:

Clear actionable insights

Lower data collection cost

Easier to explain to management

Better for new market entry



Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer:

Excluded (Original Top 5): 1. GrLivArea 2. OverallQual 3. TotalBsmtSF 4. GarageArea 5. YearBuilt

New Top 5:

TotalSF = TotalBsmtSF + 1stFlrSF + 2ndFlrSF

Neighborhood_NridgHt (premium location)

ExterQual_Ex (exterior quality = excellent)

KitchenQual_Ex (kitchen quality = excellent)

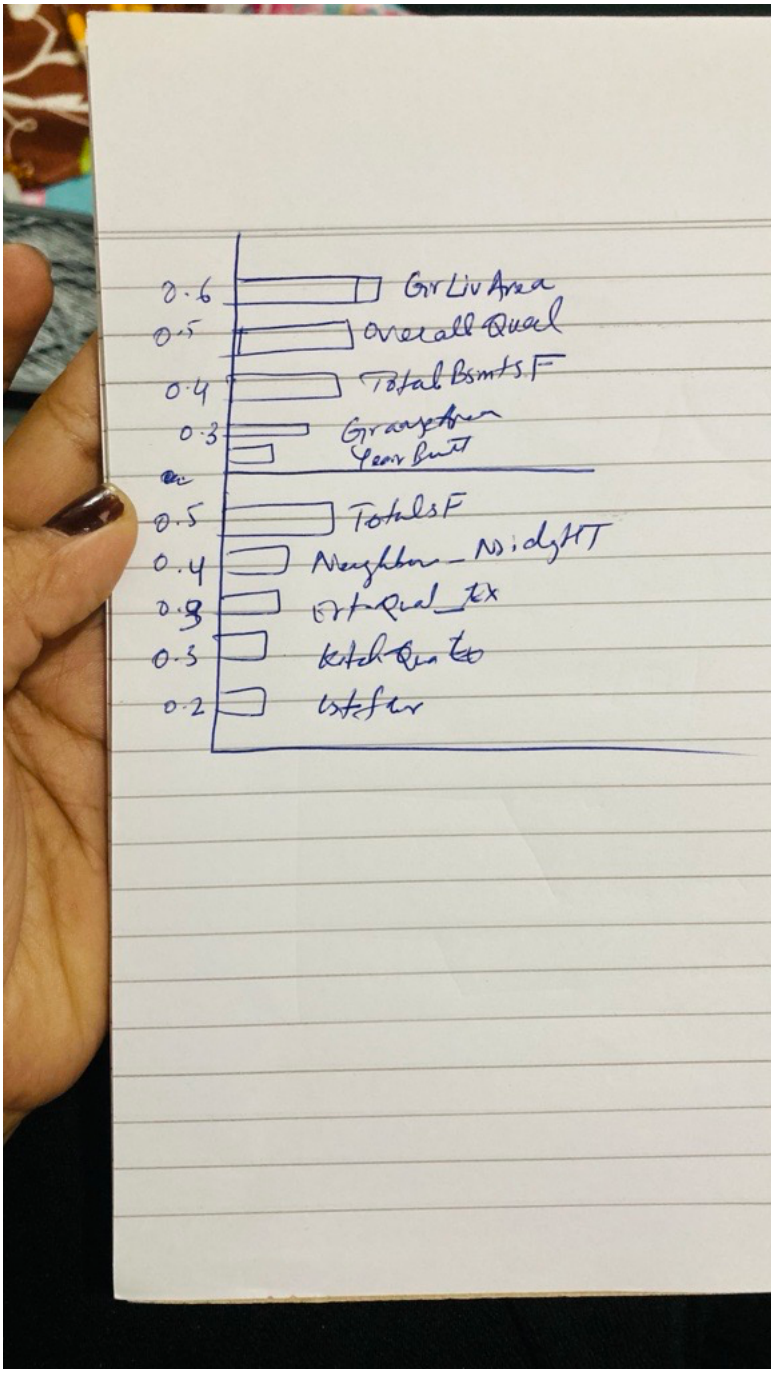
1stFlrSF (first floor sq ft)

Performance:

Original: R^2 = 0.88

New model: R^2 = 0.76

Loss: ΔR^2 = -0.12 (12% decrease



Question 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answer:

Robustness Techniques:

k-fold Cross-Validation

Split data: k parts

Train: k-1 folds

Validate: 1 fold

Repeat k times

Train-Test Split

Train: 70%

Test: 30%

Rule: Never touch test during training

Regularization

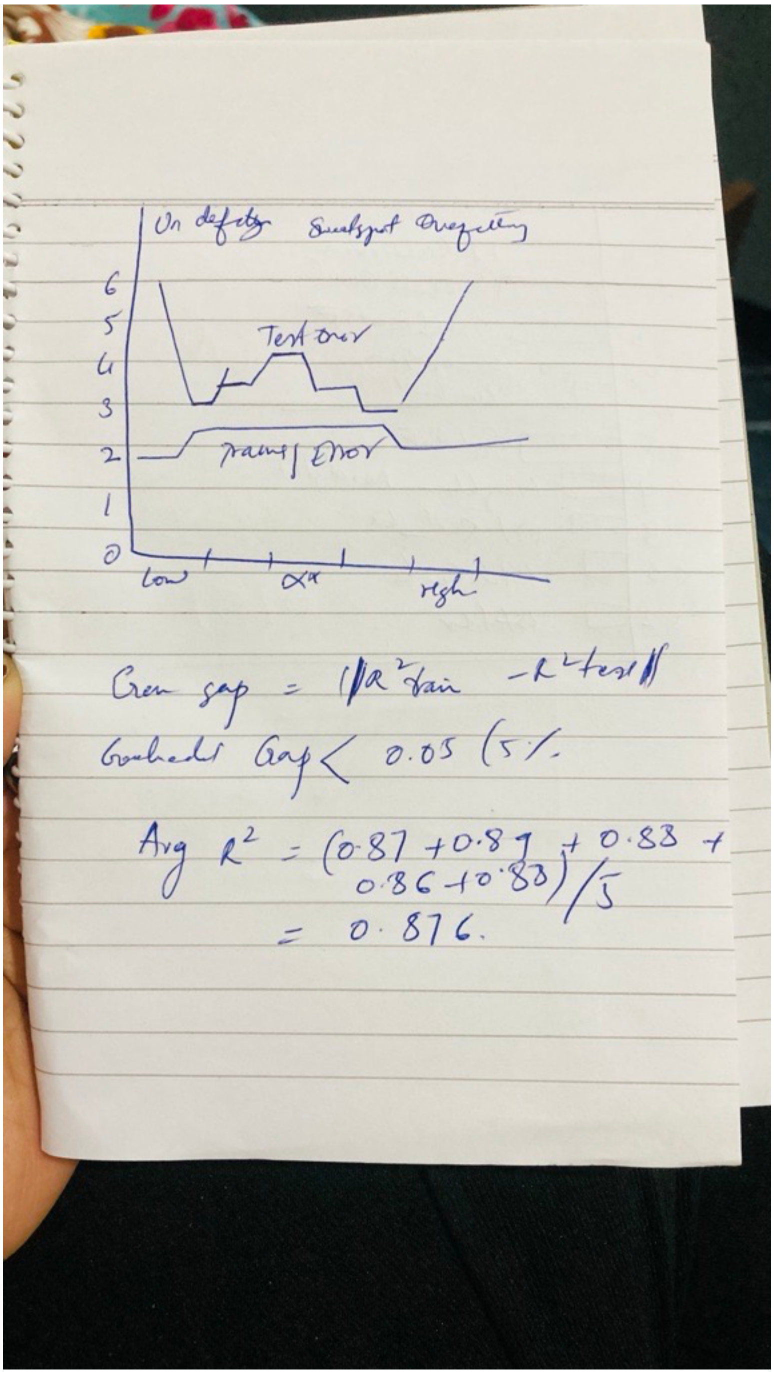
L1 (Lasso): $\min[MSE + \alpha \Sigma|\theta|]$

L2 (Ridge): $\min[MSE + \alpha \Sigma\theta^2]$

Bias-Variance Tradeoff:

Total Error = $Bias^2 + Variance + \varepsilon$

Where: ε = irreducible error



Impact on Accuracy:

Scenario	Train R^2	Test R^2	Gap	Status
Overfitted	0.98	0.72	26%	❌ Bad
Robust	0.89	0.87	2%	✓ Good

Why Robust ⇒ Lower Train Accuracy:

Regularization penalizes large θ

Ignores noise patterns

Simpler model can't fit all points

BUT: Better on new data