Why did the Linear SVM fail, and why did the RBF kernel succeed?

The linear SVM failed because the *moons* dataset is non-linearly separable.
In the scatter plot and linear SVM decision boundary, you can clearly see:

- The data forms two moon-shaped curves.
- There is no single straight line (linear hyperplane) that can separate the two classes correctly.
- As a result, the linear model misclassifies many points, giving comparatively low accuracy and messy boundary.

In contrast, the RBF kernel succeeded because:

- RBF uses the kernel trick to map the data into a higher-dimensional space.
- In that transformed space, the moons become linearly separable.
- The RBF kernel draws a smooth, curved boundary that wraps around the moon shape.
- This dramatically improves classification accuracy, clearly visible in the RBF plot where almost every point is correctly separated.

What did GridSearchCV tell you? What were the best C and gamma?

GridSearchCV performs a systematic search over combinations of hyperparameters and picks the model that gives the highest cross-validated accuracy.

- C = 10 → the model allows fewer misclassifications (higher penalty), giving a clearer and stricter boundary.
- Gamma = 1 → influences how far the effect of a single data point spreads.
    - Too small gamma → boundary too smooth, underfits.
    - Too large gamma → boundary too sharp, overfits.

So, GridSearchCV finds the sweet spot: a boundary flexible enough to capture patterns but not overly complex.

What happens if gamma is too high (e.g., 1000)?

If gamma is extremely high:

- The influence of each training point becomes very small and localized.
- The model tries to fit every point exactly.
- Decision boundary becomes very wiggly and extremely complex.
- The boundary overfits noise and looks like it draws tight loops around individual points.

✅ Effect:
✔ Training accuracy very high

❌ Validation accuracy decreases
❌ Decision boundary looks irregular and overfitted

What happens if C is too low (e.g., 0.01)?

C controls the penalty for misclassification.

- When C is small, the model allows many misclassifications to keep the boundary simple.
- The decision boundary becomes too smooth and broad, not fitting the moons properly.
- Underfitting occurs.

✅ Effect:
❌ Many errors allowed
❌ Boundary too simple
❌ Accuracy drops
✔ Model becomes very generalized but not accurate