a. Best model based on accuracy:

From my results:

- **Linear Regression**: 0.96
- **Polynomial Regression (degree 2)**: 0.9733
- **Polynomial Regression (degree 3)**: 0.9467
- **Naive Bayes (GaussianNB)**: 0.9667
- **kNN (k=5)**: 0.9467
- **LDA**: 0.98
- **QDA**: 0.9733

Thus, the best model is **Linear Discriminant Analysis (LDA)** with an accuracy of **0.98**.

b. Why other models perform slightly worse

1. **Linear Regression (0.96)**
   - Regression is not designed for classification; it predicts continuous values that must be rounded.
   - This rounding may misclassify borderline samples.
   - It does not model class boundaries directly, so it underperforms compared to LDA/QDA.
2. **Polynomial Regression (degree 2) (0.9733)**
   - Very close to QDA and almost as good as LDA.
   - However, polynomial regression can **overfit** on training folds, which may reduce its ability to generalize perfectly to the test fold.
3. **Polynomial Regression (degree 3) (0.9467)**
   - Higher degree polynomials introduce **too much flexibility**.
   - The model captures noise and overfits, which hurts generalization.

- Accuracy drops because the class decision boundaries become too complex.

4. **Naive Bayes (0.9667)**
   - Assumes **feature independence** given the class, which isn't true for the iris dataset (e.g., petal length and width are correlated).
   - Still performs well, but not as high as LDA which models covariance between features.
5. **kNN (0.9467)**
   - Relies on local neighborhoods.
   - Sensitive to class overlap and the chosen value of **k**.
   - With k=5, some misclassifications happen when neighbors belong to multiple classes, especially between versicolor and virginica (which are harder to separate).
6. **QDA (0.9733)**
   - Slightly worse than LDA.
   - QDA allows each class to have its own covariance matrix, which increases flexibility.
   - This flexibility can backfire on small datasets like Iris (150 samples), where covariance estimates may not be stable → leading to minor misclassifications.