# MA477: Data Science

**Lesson 17 Outline — 24 February 2026**

United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

# 1 Administrative

- Student review
- Work Problem Set 2 due tonight
- Classification comparison discussion

| **Method** | **Bias–Var** | **Main Params** | $n$ | $p$ | **Comp.** |
|---|---|---|---|---|---|
| Logistic Reg | ↑Bias ↓Var (linear) | $\lambda$ or $C$, L1/L2 | OK small–med; ↑ $n$ helps | Handles med/high $p$ (reg) | Fast (iterative) |
| LDA | ↑Bias ↓Var (if Gaussian, $\Sigma_k = \Sigma$) | Priors; $\Sigma$ est. | Strong small $n$ (if assumptions) | Weak if $p \approx n$ (cov est.) | Very fast |
| SVM (Hard) | Low Bias; Var ↑ if noisy | – (sep. req.) | Needs separable data | Works high $p$ (linear) | Med |
| SVM (Soft) | $C$ ↑: Bias↓, Var↑ | $C$ | Good small–med $n$ | Linear OK high $p$ | Med |
| SVM (Kernel) | Flex ↑ Bias↓, Var↑ | Kernel, $C$, $\gamma$ | Med $n$; large $n$ costly | Handles nonlinear; risk high $p$ small $n$ | High (kernel matrix) |
| KNN | $k$ ↑: Bias↑, Var↓ | $k$, metric; scale! | Needs larger $n$ | Poor high $p$ (curse) | Train cheap; pred costly |

Table 1: Compact comparison: $n$ = sample size, $p$ = dimension. Arrows indicate relative increase/decrease.

# 2 Classification Comparison Lesson Objectives

- Review all previous classification lessons since lesson 9

# Concept Review Questions with Short Answers

1. **Soft-margin SVM and $C$:** In a soft-margin SVM, how does $C$ control the tradeoff between a wide margin and margin violations? What happens to bias and variance as $C$ changes?

   **Answer:** Large $C$ heavily penalizes violations, leading to a narrower margin, lower bias, and higher variance. Small $C$ allows more violations, giving a wider margin, higher bias, and lower variance.

2. **Maximum-margin idea:** What does it mean to maximize the margin, and why can this improve generalization?

   **Answer:** Maximizing the margin means maximizing the distance from the boundary to the closest training points. Larger margins tend to reduce variance and improve generalization.

3. **Kernel purpose:** What is the main purpose of a kernel function in SVM?

   **Answer:** A kernel allows SVM to fit nonlinear boundaries by implicitly mapping data into a higher-dimensional space without explicitly computing the transformation.

4. **RBF kernel and $\gamma$:** What does $\gamma$ control in an RBF SVM?

   **Answer:** $\gamma$ controls how local each point's influence is. Large $\gamma$ gives very flexible, wiggly boundaries (low bias, high variance). Small $\gamma$ gives smoother boundaries (high bias, low variance).

5. **Feature scaling:** Why is scaling important for some models? Name two examples.

   **Answer:** Scaling ensures features contribute equally to distance or similarity calculations. It is especially important for KNN and SVM.

6. **KNN and $k$:** How does changing $k$ affect bias and variance?

   **Answer:** At $k = 1$, KNN has low bias and high variance. As $k$ increases, the boundary becomes smoother, increasing bias and reducing variance.

7. **Curse of dimensionality:** Why does KNN struggle in high dimensions?

   **Answer:** In high dimensions, distances between points become similar, making nearest neighbors less meaningful and reducing predictive power.

8. **Logistic regression vs SVM:** What is the key conceptual difference?

   **Answer:** Logistic regression models class probabilities directly, while SVM focuses on maximizing the margin. Logistic regression is generally more interpretable.

| **Parameter** | **Where (SVC)** | **Effect if Increased** | **Bias–Variance Interpretation (Concise)** |
|---|---|---|---|
| $C$ | All kernels | Fewer training errors; narrower margin; more support vectors can become "active" | **Variance up, bias down.** Larger $C$ fits training data more tightly (risk of overfitting). Smaller $C$ allows more violations (more regularization). |
| `kernel` | Chooses model family (e.g., `linear`, `rbf`, `poly`) | More flexible kernels can fit more complex boundaries | **Flexibility affects bias–variance.** Linear tends to higher bias / lower variance; RBF and higher-degree polynomial tend to lower bias / higher variance. |
| $\gamma$ | RBF, Polynomial, Sigmoid | More local / sharper influence of points (RBF); stronger scaling of dot products (Poly) | **Variance up, bias down.** Large $\gamma$ yields very flexible boundaries (overfitting risk). Small $\gamma$ yields smoother, more global boundaries (underfitting risk). |
| `degree` ($d$) | Polynomial kernel only | More complex polynomial interactions; more curvature | **Variance up, bias down.** Higher degree increases model complexity quickly; lower degree is smoother and more biased. |
| `coef0` ($r$) | Polynomial (also Sigmoid) | In polynomial kernels, increases the influence of lower-order terms relative to pure high-order interactions | **Often bias up, variance down (but data-dependent).** Larger `coef0` typically makes the kernel behave less "purely high-degree," which can stabilize fits; very small `coef0` can emphasize high-order interactions and increase variance. |
| `gamma = "scale"` | Default choice for $\gamma$ (RBF/Poly) | Sets $\gamma$ based on data variance and number of features | **Stabilizes variance across datasets.** Helps prevent extreme $\gamma$ values when features are not standardized; still may need tuning. |
| `gamma = "auto"` | Alternative choice for $\gamma$ (RBF/Poly) | Sets $\gamma$ to $1/p$ (p = number of features) | **Simpler, can miscalibrate complexity.** May be too flexible or too smooth depending on feature scaling; often less robust than `"scale"`. |

Table 2: SVM (scikit-learn `SVC`) tuning parameters and their bias–variance interpretations.