Flexible vs. Inflexible Methods

---

1. **Flexible Methods**

Definition:

- Methods that make **fewer assumptions** about the functional form of the relationship between predictors and response.
- They try to let the data "speak for itself."

Examples:

- k-Nearest Neighbors (kNN)
- Decision Trees, Random Forests
- Splines, Generalized Additive Models (GAMs)
- Neural Networks, Deep Learning

Strengths:

- Can capture **complex, nonlinear relationships**.
- Perform well when:
  - Sample size n is large (reduces variance).
  - A true relationship is highly nonlinear.

Weaknesses:

- High variance: prone to **overfitting**, especially if n is small or the data is noisy.
- Computationally more expensive.
- Harder to interpret.

---

2. **Inflexible (Restrictive) Methods**

Definition:

- Methods that make **strong assumptions** about the functional form (e.g., linear or simple structure).
- They impose structure even if the data suggests otherwise.

Examples:

- Linear Regression
- Logistic Regression
- Linear Discriminant Analysis (LDA)
- Naïve Bayes

Strengths:

- Lower variance, less risk of overfitting.
- Works better with:
  - Small sample size n.
  - Many predictors **p** relative to **n**.
  - High noise environments.
- Easier to interpret.

Weaknesses:

- Can **underfit** if the true relationship is nonlinear or complex.
- Less flexible in capturing interactions automatically.

---

3. **Key Tradeoff: Bias–Variance**

- **Flexible methods** → Low bias, high variance.
- **Inflexible methods** → High bias, low variance.

Choosing between them depends on:

- Size of dataset (n).
- Number of predictors (p).
- Complexity of true relationship.

- Noise level (σ^2).

✅ **Rule of Thumb:**

- Use **flexible methods** when you have **lots of data** and expect **nonlinear patterns**.
- Use **inflexible methods** when data is **limited**, predictors are **many**, or noise is **high**.

Aspect	**Flexible Methods**	**Inflexible Methods**
Definition	Make few assumptions about data structure	Make strong assumptions about data structure
Examples	kNN, Decision Trees, Random Forests, Splines, Neural Networks	Linear Regression, Logistic Regression, LDA, Naïve Bayes
Bias	Low bias (can fit complex patterns)	High bias (restricted to simple forms)
Variance	High variance (risk of overfitting)	Low variance (less sensitive to noise)
Data Requirement	Need large n (many observations) to work well	Can work with small n (fewer observations)
Predictors (pp)	Struggle if p is large relative to nn (curse of dimensionality)	Handle large p better if assumptions hold
Noise Handling	Overfits when noise is high	More robust to noise
Relationship	Capture nonlinear & complex relationships	Perform well only if relationship is simple/linear
Interpretability	Harder to interpret	Easier to interpret
Computation	More computationally expensive	Usually computationally cheaper
When to Use	- Large data size (n) - Relationship is complex & nonlinear	- Small sample size - High noise - Relationship is approximately linear