Linear Models

What Are Linear Models?

- Predict outcomes using a **weighted sum of attributes**
- Two major tasks:
  - **Numeric prediction** $\rightarrow$ linear regression
  - **Classification** $\rightarrow$ logistic regression, perceptron, Winnow
- Simple yet powerful; used throughout ML and statistics

General Linear Model

$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k$$

Where: - x = predicted value
- w_0 = bias term
- w_i = weights learned from data

Linear Regression

- Used for **numeric prediction**.
- Goal: minimize squared error:

$$\sum (actual - predicted)^2$$

- Uses least-squares estimation

Linear Regression Steps

1. Gather training examples
2. Build attribute matrix
3. Add bias column
4. Solve for weight vector:

$$w = (A^T A)^{-1} A^T x$$

5. Use w to predict new values.

Linear Regression: When It Works

- Best for:
  - Linear or near-linear relationships
  - Numeric features
  - Interpretability needed
- Limitations:
  - Sensitive to outliers
  - Cannot model nonlinear interactions without feature engineering

Logistic Regression

- Used for **binary classification**.
- Outputs a **probability**:

$$P(y = 1 \mid a) = 1/(1 + e^{-(w_0 + w_1 a_1 + \ldots + w_k a_k)})$$

Why Logistic Instead of Linear Regression?

- Linear regression can output values <0 or >1
- Logistic uses the **sigmoid** to bound outputs
- Uses **maximum likelihood**, not least squares

Logistic Regression Boundary

- Decision rule:

  ```
  Predict class 1 if P >= 0.5
  ```

- Boundary is where:

  ```
  w0 + w1*a1 + ... + wk*ak = 0
  ```

- This forms a **hyperplane**.

The Perceptron

- A simple linear classifier.
- Prediction:

```
if (w * a > 0) -> class = 1
else -> class = -1
```

Perceptron Update Rule

- If positive example misclassified:

  ```
  w = w + a
  ```

- If negative misclassified:

  ```
  w = w - a
  ```

- Converges if data is linearly separable.

Perceptron Properties

- Fast and online
- Foundation of neural networks
- Only works for **linearly separable** data

The Winnow Algorithm

- Designed for:
  - **Binary attributes**
- High-dimensional data with few relevant features (attribute-efficient)
  - Uses **multiplicative** weight updates.

Winnow Update Rule

- If prediction incorrect:
  - Positive example: multiply weights where attribute=1 by α
  - Negative example: divide weights where attribute=1 by α
- Weights remain non-negative.

Balanced Winnow

- Uses two weight vectors:

  `w+ for positive evidence`
  `w- for negative evidence`

- Predict positive if:

  `w+ * a - w- * a > threshold`

- Allows negative evidence handling.

Strengths of Linear Models

- Fast and scalable
- Simple and interpretable
- Form basis of many advanced models (NNs, SVMs)
- Often strong baselines

Limitations

- Only models linear boundaries
- Sensitive to outliers and correlated attributes
- Cannot model interactions without engineered features

Summary

- Linear regression -> numeric prediction
- Logistic regression -> probability-based classification
- Perceptron -> simple linear classifier
- Winnow -> efficient for sparse binary data
- Linear models = simple, powerful, widely used.