1 Problem Formulation

Input Space

$$\mathcal{X} \subseteq \mathbb{R}^p$$

Output Space

$$\mathcal{Y} = \{0, 1\}$$

Model

$$P(Y = 1|x) = \sigma(x^T \beta)$$

where:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

2 Model Specification

Hypothesis

$$h_\beta(x) = \sigma(x^T \beta)$$

Structural Assumption

- Log-odds linear:

$$\log \frac{p}{1 - p} = x^T \beta$$

3 Loss Function

Logistic Loss

$$L = -y \log p - (1 - y) \log(1 - p)$$

Convexity Proof

Hessian:

$$\nabla^2 = X^T W X$$

where W diagonal with positive entries.

Hence positive semi-definite → convex.

4 Objective Function

$$J(\beta) = -\sum_{i=1}^{n} [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$

Regularized:

$$J(\beta) + \lambda ||\beta||^2$$

---

5 Optimization

No Closed Form

Iterative methods:

- Gradient Descent
- Newton's Method

IRLS (From Newton)

Update:

$$\beta^{new} = (X^T W X)^{-1} X^T W z$$

6 Statistical Interpretation

Bernoulli Model

$$Y \sim Bernoulli(p)$$

Logistic regression = MLE for Bernoulli likelihood.

Maximum Entropy View

Logistic regression is the maximum entropy classifier subject to linear constraints.

---

7 Regularization & Generalization

Less prone to overfitting than high-degree polynomial regression because:

- Convex loss
- Linear decision boundary
- Capacity limited

8 Theoretical Properties

Perfect Separation

If data perfectly separable:

$$||\beta|| \to \infty$$

MLE does not exist (likelihood keeps increasing).

Asymptotic Distribution

$$\hat{\beta} \sim N(\beta, (X^T W X)^{-1})$$

Logistic vs Hinge

Logistic	Hinge
Smooth	Non-smooth
Probabilistic	Margin-based
Convex	Convex

Generative vs Discriminative

Logistic	Naive Bayes
Models (P(Y	X))
Fewer assumptions	Strong independence assumption
Lower bias	Higher bias

9 Computational Complexity

Stage	Complexity
Training	$O(knp^2)$
Inference	$O(p)$
Memory	$O(np)$

---

10 Limitations

- Fails under perfect separation
- Sensitive to extreme outliers
- Linear boundary assumption
- Struggles with highly nonlinear data