

# Tutorial: Logistic Regression

## *Cheat Sheet and Practice Problems*

ES335 - Machine Learning  
IIT Gandhinagar

July 23, 2025

## 1 Summary from Slides

### 1.1 Problem Setup

**Classification Task:** Predict discrete/categorical outputs (e.g., Orange vs Tomato)

**Why not Linear Regression?**

- Linear regression outputs continuous values
- Classification needs probabilities between 0 and 1
- Linear regression can output values outside  $[0,1]$  range

### 1.2 Logistic/Sigmoid Function

**Sigmoid Function:**

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

**Properties:**

- Maps any real number to  $(0,1)$
- $\sigma(0) = 0.5$
- $\sigma(\infty) = 1, \sigma(-\infty) = 0$
- S-shaped curve

**Logistic Regression Model:**

$$P(y = 1|x) = \sigma(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

where  $\theta^T x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$

### 1.3 Maximum Likelihood Estimation

**Likelihood Function:**

$$L(\theta) = \prod_{i=1}^m P(y^{(i)}|x^{(i)}; \theta)$$

**For Binary Classification:**

$$P(y|x; \theta) = h_{\theta}(x)^y (1 - h_{\theta}(x))^{1-y}$$

where  $h_{\theta}(x) = \sigma(\theta^T x)$

**Log-Likelihood:**

$$\ell(\theta) = \sum_{i=1}^m [y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

## 1.4 Cross-Entropy Cost Function

**Cost Function** (negative log-likelihood):

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

**Intuition:**

- When  $y = 1$ : Cost =  $-\log(h_{\theta}(x))$  (penalizes low probability predictions)
- When  $y = 0$ : Cost =  $-\log(1 - h_{\theta}(x))$  (penalizes high probability predictions)
- Convex function - single global minimum

**Gradient:**

$$\frac{\partial J}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

## 2 Practice Problems

**Problem : Sigmoid Function Properties**

a) Calculate  $\sigma(0)$ ,  $\sigma(2)$ , and  $\sigma(-3)$  b) Show that  $\sigma(-z) = 1 - \sigma(z)$  c) Prove that  $\frac{d\sigma(z)}{dz} = \sigma(z)(1 - \sigma(z))$

**Problem : Binary Classification Setup**

For a simple 1D logistic regression  $P(y = 1|x) = \sigma(\theta_0 + \theta_1 x)$  with parameters  $\theta_0 = -2, \theta_1 = 1$ :

a) What is the decision boundary? b) Calculate  $P(y = 1|x = 3)$  c) At what value of  $x$  is the probability exactly 0.8?

**Problem : Cost Function Calculation**

Given one training example  $(x = 2, y = 1)$  and model parameters  $\theta_0 = 0.5, \theta_1 = -0.3$ :

a) Calculate the predicted probability  $h_{\theta}(x)$  b) Calculate the cost for this single example c) What would the cost be if  $y = 0$  instead?

**Problem : Gradient Computation**

For the dataset:

$x$	$y$
1	0
2	0
3	1
4	1

With current parameters  $\theta_0 = 0, \theta_1 = 0$ :

a) Calculate predictions for all examples b) Compute the gradient  $\frac{\partial J}{\partial \theta_0}$  and  $\frac{\partial J}{\partial \theta_1}$  c) In which direction should parameters move to reduce cost?

**Problem : Decision Boundary**

For 2D logistic regression  $P(y = 1|x_1, x_2) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$ :

- a) Derive the equation of the decision boundary b) If  $\theta = [-1, 2, 3]^T$ , what is the decision boundary equation? c) Classify the points (1, 1), (0, 1), and (2, 0)

**Problem : Maximum Likelihood Derivation**

For a single training example  $(x, y)$ :

- a) Write the likelihood  $P(y|x; \theta)$  for logistic regression b) Take the log and derive the log-likelihood c) Show how maximizing log-likelihood leads to minimizing cross-entropy loss

**Problem : Multiclass Extension**

Extend binary logistic regression to 3-class classification:

- a) Write the softmax function for 3 classes b) What is the decision rule for classification? c) How many parameters are needed for 3 classes with 2 features? d) Write the cross-entropy loss for multiclass case

**Problem : Regularization**

Add L2 regularization to logistic regression:

- a) Write the regularized cost function b) Derive the regularized gradient c) How does regularization affect the decision boundary? d) What happens when  $\lambda \rightarrow \infty$ ?

**Problem : Newton's Method**

For logistic regression, Newton's method update is:

$$\theta := \theta - H^{-1} \nabla J(\theta)$$

- a) What is the Hessian matrix  $H$  for logistic regression? b) Why is Newton's method faster than gradient descent for logistic regression? c) What are the computational trade-offs?

**Problem : Feature Engineering**

You have categorical features for predicting email spam:

- a) How would you encode a categorical feature "Domain" with values gmail, yahoo, hotmail? b) What about interaction terms between "Time of Day" and "Day of Week"? c) How would you handle polynomial features in logistic regression?

**Problem : Model Evaluation**

Your logistic regression model outputs these probabilities and true labels: Predictions: [0.9, 0.3, 0.8, 0.1, 0.7] True labels: [1, 0, 1, 0, 1]

- a) Calculate accuracy using 0.5 threshold b) Calculate log-loss (cross-entropy) c) How would you choose the optimal threshold? d) What if the dataset is imbalanced (90% negative class)?

**Problem : Convergence Analysis**

During gradient descent training, you observe: - Iteration 1: Cost = 0.8 - Iteration 10: Cost = 0.5 - Iteration 100: Cost = 0.49 - Iteration 1000: Cost = 0.489

a) Is the algorithm converging properly? b) What might cause slow convergence? c) How would you diagnose overfitting vs underfitting? d) When should you stop training?

**Problem : Comparison with Linear Regression**

Compare logistic and linear regression for classification:

a) What happens if you use linear regression for binary classification? b) Why is cross-entropy loss preferred over MSE for classification? c) Can you use R-squared for logistic regression evaluation? d) How do the decision boundaries differ?

**Problem : Real-world Application**

Design a logistic regression model for medical diagnosis:

a) You have features: age, blood pressure, cholesterol, family history. Write the model. b) How would you handle missing values in features? c) What ethical considerations arise from false positives vs false negatives? d) How would you validate the model for clinical use? e) What if you need to explain predictions to doctors?

**Problem : Advanced Challenge**

Implement coordinate ascent for logistic regression:

a) Derive the update rule for one parameter while keeping others fixed b) Why might coordinate ascent be preferred over gradient ascent in some cases? c) How does the algorithm handle non-differentiable regularizers like L1? d) Compare convergence properties with batch and stochastic gradient descent