1 Introduction to ML and DS Formulas

Machine learning (ML) and data science (DS) rely on mathematical foundations to model and analyze data. This document presents key formulas used in these fields, formatted to span three pages.

2 Linear Regression

Linear regression models the relationship between a dependent variable y and independent variable(s) x using a linear equation:

$$y = \beta_0 + \beta_1 x + \epsilon$$

where:

- β_0: Intercept.
- β_1: Slope coefficient.
- ϵ: Error term.

The least squares method minimizes the cost function:

$$J(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

where n is the number of observations.

To fit the model, the normal equations are solved:

$$\beta = (X^T X)^{-1} X^T y$$

where X is the design matrix and y is the response vector.

3 Logistic Regression

Logistic regression predicts binary outcomes using the sigmoid function:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

The log-odds (logit) is:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

The cost function (log-loss) to optimize is:

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

4 Gradient Descent

Gradient descent optimizes the cost function by iteratively adjusting parameters:

$$\theta_j := \theta_j - \alpha\frac{\partial J(\theta)}{\partial \theta_j}$$

where:

- θ_j: Parameter to update.
- α: Learning rate.
- $\frac{\partial J(\theta)}{\partial \theta_j}$: Gradient of the cost function.

For linear regression, the gradient is:

$$\frac{\partial J}{\partial \theta_j} = \frac{2}{n}\sum_{i=1}^{n}(h_\theta(x_i) - y_i)x_{i,j}$$

where $h_\theta(x) = \theta^T x$ is the hypothesis.

5 Entropy and Information Gain

Entropy measures uncertainty in a dataset:

$$H(S) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

where p_i is the probability of class i.

Information gain for a split is:

$$\mathrm{IG}(S, A) = H(S) - \sum_{v \in \mathrm{Values}(A)} \frac{|S_v|}{|S|} H(S_v)$$

where S_v is the subset of S for attribute A value v.

6 Additional ML/DS Notes

Common distance metrics include Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

and cosine similarity:

$$\mathrm{cosine}(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

These are widely used in clustering and recommendation systems. For large datasets, regularization (e.g., L2: $\lambda \|\beta\|^2$) prevents overfitting in regression models.