

Applications - Lecture 5

Maximum Likelihood Estimation

Ingénierie Mathématique & Informatique

26/10/2026

Section	Notion	Used in
5.1	Likelihood, MLE definition	App. 1
5.2	Fisher information, Cramer-Rao, asymptotic efficiency	App. 1
5.3.1	Logistic regression likelihood	App. 1
5.3.2	Mixture models, EM algorithm	

1. **Classify**: binary labels via logistic regression, MLE seen as cross-entropy minimisation.

App. 1 - Logistic Regression and Cross-Entropy Loss

From likelihood to cross-entropy

Problem. Given pairs $(x_i, y_i) \in \mathbb{R}^p \times \{0, 1\}$, estimate the probability that a new input x belongs to class 1.

Model. The Y_i are independent Bernoulli with

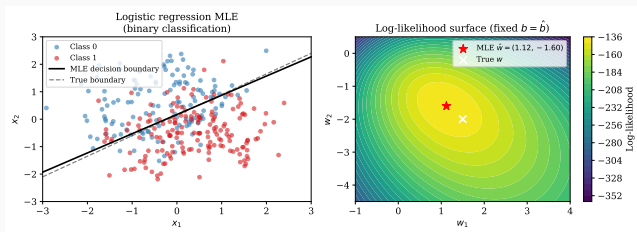
$$p_\beta(x_i) = \Psi(\beta_0 + \beta_1 x_i^1 + \cdots + \beta_p x_i^p), \quad \Psi(u) = \frac{1}{1 + e^{-u}},$$

so the likelihood of (x_n, y_n) is (extended to a non-iid sample)

$$L_n(x_n, y_n; \beta) = \prod_{i=1}^n p_\beta(x_i)^{y_i} (1 - p_\beta(x_i))^{1-y_i}.$$

The quantity $-\ell_n(x_n, y_n; \beta)$ is exactly the **cross-entropy** loss: minimising cross-entropy in machine learning *is* maximum likelihood estimation. The MLE $\hat{\beta}_n$ has no closed form here and is found by numerical solvers.

Efficiency of the MLE and take-aways



Left: MLE boundary (solid) vs. true boundary (dashed). Right: log-likelihood contours in (w_1, w_2) , MLE (star) vs. true parameter (cross).

- The estimator $\hat{\beta}_n$ is **asymptotically efficient**: $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, I(\beta)^{-1})$
- Contour curvature (right panel) is the Fisher information $I(\beta)$: elongated ellipses signal poorly determined directions.
- The boundary is linear in x : a linear classifier trained by negative log-likelihood.