Foundations of Machine Learning
AI2000 and AI5000

FoML-11
Bayesian Regression

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

So far in FoML

- What is ML and the learning paradigms
- Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Linear Regression with basis functions - and regularization
- Model selection
- Bias-Variance Decomposition/Tradeoff (Bayesian Regression)

Decision Theory

Decision Theory

- Dataset: i/p vectors $\mathbf{x} \in \mathbb{R}^D$, ground truth $t \in \{C_1, C_2, \ldots, C_K\}$
- Divide the i/p space $\mathbb{R}^D$ into K decision regions $R_k,\ k = \{1, 2, \ldots K\}$
- For every data point
  - Ground truth
  - Prediction

Decision Theory

- Confusion Matrix

Diagonal elements -

Off-diagonal elements -

Decision Theory - Misclassification Rate

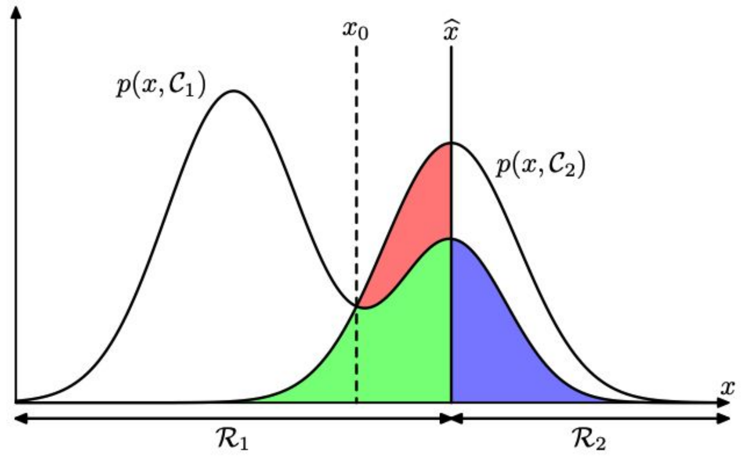
- Goal of classification - Minimize the misclassification rate
- Assume the data are drawn independently from the joint distribution
- Probability of a misclassification: $p(\text{mistake}) = \sum_{i=1}^{K} \sum_{k \neq i} p(\mathbf{x} \in R_i, C_k)$

$$p(\text{mistake}) =$$

Decision Theory - Misclassification Rate

- Minimizing the misclassification rate
  - Assign x to class C_k if $p(\mathbf{x}, t = C_k) > p(\mathbf{x}, t = C_j), \;\; \forall j \neq k$
  - We know that

Decision Theory - Misclassification Rate



Minimizing the Misclassification Rate - Issues

- Not all errors have the same impact!
- E.g. medical diagnosis
  - E1:
  - E2:

Minimizing the Misclassification Rate - Issues

- Class imbalance
  - May lead to skewed view of the classifier's performance

Expected Loss

- Possible solution: use different weights for different error types

$$\mathrm{L} = \begin{pmatrix} 0 & \\ & 0 \end{pmatrix}$$

$$\mathbb{E}[L] = \sum_{k,j} L_{k,j} \int_{\mathcal{R}_j} p(x, C_k) dx$$

Minimize the expected loss: (assign x to Ck if) $\sum_{j=1}^{K}$ is minimal

Classification Strategies

- Discriminant functions
  - Direct functions of i/p to target $t = y(\mathbf{x}, \mathbf{w})$
- Probabilistic Discriminant models
  - Posterior class probabilities $p(C_k/\mathbf{x})$
- Probabilistic generative models
  - Class-conditional models $p(\mathbf{x}/C_k)$
  - Prior class probabilities $p(C_k)$

Next
Probabilistic Generative Models