|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phương pháp | Input | Output | Bài toán ưu (hoặc hàm mất mát) | Cách thực hiện |
| Linear Regression (Hồi quy tuyến tính) |  |  |  |  |
| Perceptron | Tập dữ liệu huấn luyện đã được gán nhãn (Xtrain, ytrain | Siêu phẳng tuyến tính f(x,w) = wTx phân chia tập dữ liệu huấn luyện sao cho các mẫu cùng nhãn thì thuộc cùng một phía của siêu phẳng |  | Dùng Gradient Dresscent:  1. Tại thời điểm t = 0, chọn ngẫu nhiên một vector hệ số w0  2. Tại thời điểm t, nếu không có điểm dữ liệu nào bị phân lớp lỗi, dừng thuật toán.  3. Giả sử xi là một điểm bị phân lớp lỗi. Cập nhật  wt+1 = wt + yi xi  4. Thay đổi t = t + 1 rồi quay lại Bước 2 |
| SVM | Tập dữ liệu huấn luyện đã được gán nhãn (Xtrain, ytrain) | Siêu phẳng f(x,w) = wTx phân chia dữ liệu huấn luyện, sao cho:   * Tất cả điểm dữ liệu có nhãn 1 thuộc về cùng một phía của siêu phẳng. * Khoảng cách từ điểm gần nhất của mỗi lớp tới siêu là bằng nhau (*margin*, *lề*). * Margin này phải là cực đại. |  | * Bài toán tối ưu này là bài toán lồi, và là một quadratic programming (phương trình bậc 2) * Suy ra nghiệm cho SVM là duy nhất * Để giải bài toán này, người ta thường giải bài toán đối ngẫu Lagrange |
| Cây quyết định |  |  |  |  |
| Neural Network | Tập dữ liệu huấn luyện, kiến trúc mạng nơ ron (số lớp ẩn, số nơ ron của mỗi lớp ẩn), hàm kích hoạt (activation function), hàm mất mát (loss function) | Bộ vector trọng số của các liên kết giữa các nơ-ron (W) để hàm mất mát đạt giá trị tối ưu | Hàm mất mát tại tầng ouput của mạng nơ-ron tùy thuộc vào bài toán thực hiện. Nếu bài toán Hồi quy, hàm mất mát có thể sử hàm mất mát của Hồi quy tuyến tính. Nếu bài toán phân lớp, có thể sử dụng hàm mất mát của Perceptron hoặc Logistic Regression, … | Phương pháp phổ biến nhất để tối ưu hàm mất mát của mạng nơ-ron là Gradient Descent (GD)  Để áp dụng GD, chúng ta cần tính được đạo hàm của hàm mất mát theo từng trọng số trong mạng nơ-ron bằng phương pháp lan truyền ngược (backpropagation) |
| Logistic Regression |  |  |  |  |
| K-means |  |  |  |  |

**ĐỀ CƯƠNG ÔN TẬP LÝ THUYẾT HỌC PHẦN HỌC MÁY** (Cô đã điền một số nội dung làm ví dụ, các em cập nhật các các nội dung còn lại)

|  |  |  |
| --- | --- | --- |
| Bài toán | Tập dữ liệu | Các phương pháp |
| Hồi quy | Tập dữ liệu đã được gán nhãn (Xtrain, ytrain). Miền giá trị của tập nhãn ytrain là miền giá trị liên tục và có vô số giá trị | Linear Regression, Lasso, Ridgre |
| Phân lớp | Tập dữ liệu đã được gán nhãn (Xtrain, ytrain). Miền giá trị của tập nhãn ytrain là rời rạc và hữu hạn | SVM, perceptron, cây quyết định, …. |
| Phân cụm | … | K-means |

|  |  |  |  |
| --- | --- | --- | --- |
| Kiểu học | Tập dữ liệu | Bài toán | Phương pháp |
| Học có giám sát | Tập dữ liệu huấn luyện được gán nhãn (Xtrain, ytrain) | Hồi quy, phân lớp | …. |
| Học không có giám sát | Tập dữ liệu huấn luyện không được gán nhãn (Xtrain) | Phân cụm | K-means |

Trình bày các phương pháp tránh overfiting, k-fond cross validation, gradient descent, Boostrapping (sinh thêm mẫu dữ liệu), các phương pháp học kết hợp (Bagging, Boosting, Stacking)