Week 00b Advanced: Supervised Learning Theory

Machine Learning for Smarter Innovation

1 Week 00b Advanced: Supervised Learning Theory

1.1 Linear Models

1.1.1 OLS Regression

Minimize: $L(w) = \|Xw - y\|^2$

Solution: $w^* = (X^T X)^{-1} X^T y$

Assumptions: - Linearity - Independence - Homoscedasticity - Normality of residuals

1.1.2 Ridge Regression

$w^* = (X^T X + \lambda I)^{-1} X^T y$

Shrinks coefficients, prevents overfitting

1.1.3 Lasso Regression

$\min_w \|Xw - y\|^2 + \lambda \|w\|_1$

Sparse solutions (feature selection)

1.2 Tree-Based Methods

1.2.1 CART Algorithm

Recursive binary splitting minimizing impurity:

Gini impurity:

$$I_G = 1 - \sum_{k=1}^{K} p_k^2$$

Entropy:

$$H = -\sum_{k=1}^{K} p_k \log_2 p_k$$

1.2.2 Random Forest

Bootstrap aggregating (bagging): - Sample data with replacement - Train tree on each sample - Average predictions

Reduces variance, prevents overfitting

1.2.3 Gradient Boosting

Sequential additive model:

$$F_m(x) = F_{m-1}(x) + \nu h_m(x)$$

where h_m fits residuals:

$$h_m =_h \sum_i L(y_i, F_{m-1}(x_i) + h(x_i))$$

1.3 SVM Theory

1.3.1 Primal Problem

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$$

Subject to: $y_i(w^T x_i + b) \geq 1 - \xi_i$

1.3.2 Dual Problem

$$\max_\alpha \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Subject to: $0 \leq \alpha_i \leq C$, $\sum_i \alpha_i y_i = 0$

1.4 Probabilistic Models

1.4.1 Logistic Regression

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

Maximum likelihood:

$$\max_w \sum_i [y_i \log \sigma(w^T x_i) + (1 - y_i) \log(1 - \sigma(w^T x_i))]$$

1.4.2 Naive Bayes

$$P(y|x) \propto P(y) \prod_i P(x_i|y)$$

Independence assumption simplifies computation

1.5 Learning Theory

1.5.1 PAC Learning

Sample complexity for ϵ-accurate, $(1 - \delta)$-confident:

$$m \geq \frac{1}{\epsilon}(\log |\mathcal{H}| + \log(1/\delta))$$

1.5.2 VC Dimension

- Linear classifiers in $\mathbb{R}^d$: $d + 1$
- Decision trees: $\Omega(n)$ for n leaves

1.6 References

- Hastie et al: Elements of Statistical Learning
- Bishop: Pattern Recognition and Machine Learning