Week 8 Tutorial Theoretical Task

Sarat Moka

LogitBoost: Gradient Boosting with Logistic Deviance

We follow exactly the notation of our generic gradient-boosting algorithm from lecture notes: At round m we have the current ensemble prediction

$$g_{m-1}(x)$$

and seek an additive update h_m and step size γ_m to minimize

$$\mathcal{L} \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathsf{Loss}\big(y_i,\; g_{m-1}(x_i) + \gamma_m\, h(x_i)\big)$$

for the *logistic deviance* (binomial loss)

$$\mathsf{Loss}\big(y,z\big) \;=\; \ln\big(1+e^{-\,y\,z}\big), \quad y_i \in \{-1,+1\}.$$

1. **Pseudo-residuals (negative gradients):**

$$r_i^{(m)} \;=\; -\frac{\partial\,\mathsf{Loss}\big(y_i,z\big)}{\partial z}\Big|_{z=g_{m-1}(x_i)} \;=\; \frac{y_i}{1+\exp\big(y_i\, g_{m-1}(x_i)\big)}.$$

2. **Second derivatives (optional):**

$$q_i^{(m)} \;=\; \frac{\partial^2\,\mathsf{Loss}\big(y_i,z\big)}{\partial z^2}\Big|_{z=g_{m-1}(x_i)} \;=\; \frac{\exp\big(y_i\, g_{m-1}(x_i)\big)}{\big(1+\exp\big(y_i\, g_{m-1}(x_i)\big)\big)^2}.$$

3. **Fit weak learner h_m:** Either

$$h_m \;=\; \arg\min_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\big(r_i^{(m)} - h(x_i)\big)^2 \quad \text{(gradient step)},$$

   or (Newton step) solve the weighted least-squares problem

$$h_m \;=\; \arg\min_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} q_i^{(m)}\left(\tfrac{r_i^{(m)}}{q_i^{(m)}} - h(x_i)\right)^2,$$

   where $\mathcal{H}$ is the family of weak learners.

4. **Line-search for γ_m:**

$$\gamma_m \;=\; \arg\min_{\gamma} \frac{1}{n}\sum_{i=1}^{n} \mathsf{Loss}\big(y_i,\; g_{m-1}(x_i) + \gamma\, h_m(x_i)\big).$$

   In the Newton-approximation one obtains the closed-form

$$\gamma_m \;=\; \frac{\sum_{i=1}^{n} r_i^{(m)}\, h_m(x_i)}{\sum_{i=1}^{n} q_i^{(m)}\, h_m(x_i)^2}.$$

5. **Update:**

$$g_m(x) \;=\; g_{m-1}(x) \;+\; \gamma_m\, h_m(x).$$

After M rounds, classification is by

$$\hat{y}(x) = \operatorname{sign}\big(g_M(x)\big), \quad \hat{p}(x) = \sigma\big(g_M(x)\big) \;=\; \frac{1}{1+e^{-\,g_M(x)}}.$$

When one uses the logistic deviance in the Newton-boosting style above, the algorithm is precisely *LogitBoost*[1].

[1] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)." The annals of statistics 28.2 (2000): 337-407.