Chapter *6* Optimization

Siheng Zhang
zhangsiheng@cvte.com

2021 年 10 月 19 日

This part corresponds to **Chapter 1,3,4 of PRML, Chapter of UML**. It mainly introduces some important properties regarding with functions: convexity, smoothness, strong convexity and Lipschitz, which are basis for the next chapters.

目录

1 from decision stump to decision tree

2 regression tree

P1

a) Note that the splitting of bucket M in T_{old} results into the two new buckets M and $M+1$ in T_{new}, hence, $R_M = \tilde{R}_M \bigcup \tilde{R}_{M+1}$, i.e., $N_M = \tilde{N}_M + \tilde{N}_{M+1}$, and the former $M-1$ buckets remain the same,

$$C_{imp}(T_{new}) = \sum_{m=1}^{M-1} N_m Q_m(T_{old}) + \tilde{N}_M \tilde{Q}_M(T_{new}) + \tilde{N}_{M+1} \tilde{Q}_{M+1}(T_{new})$$

So,

$$\Delta = C_{imp}(T_{old}) - C_{imp}(T_{new}) = N_M Q_M(T_{old}) - \tilde{N}_M \tilde{Q}_M(T_{new}) - \tilde{N}_{M+1} \tilde{Q}_{M+1}(T_{new})$$
$$= \sum_{i:x_i \in R_M} \left(y_i - \frac{1}{N_m} \sum_{i:x_i \in R_M} y_i \right)^2 - \sum_{i:x_i \in \tilde{R}_M} \left(y_i - \frac{1}{\tilde{N}_M} \sum_{i:x_i \in \tilde{R}_M} y_i \right)^2 - \sum_{i:x_i \in \tilde{R}_{M+1}} \left(y_i - \frac{1}{\tilde{N}_{M+1}} \sum_{i:x_i \in \tilde{R}_{M+1}} y_i \right)^2$$

b) from a), we know that $R_M = \tilde{R}_M \bigcup \tilde{R}_{M+1}$, so

$$\sum_{i:x_i \in \tilde{R}_M} \left(y_i - \frac{1}{\tilde{N}_{M+1}} \sum_{i:x_i \in \tilde{R}_M} y_i \right)^2 \leq \sum_{i:x_i \in \tilde{R}_M} \left(y_i - \frac{1}{N_m} \sum_{i:x_i \in R_M} y_i \right)^2$$

$$\sum_{i:x_i \in \tilde{R}_{M+1}} \left(y_i - \frac{1}{\tilde{N}_{M+1}} \sum_{i:x_i \in \tilde{R}_{M+1}} y_i \right)^2 \leq \sum_{i:x_i \in \tilde{R}_{M+1}} \left(y_i - \frac{1}{N_m} \sum_{i:x_i \in R_M} y_i \right)^2$$

And the right-side of two inequalities sum up to $\sum_{i:x_i \in R_M} \left(y_i - \frac{1}{N_m} \sum_{i:x_i \in R_M} y_i \right)^2$, which leads to $\Delta \geq 0$.

c) Note that $|T_{old}| = M, |T_{new}| = M+1$. Since $R^2 = 1 - \frac{C_{imp}(T)}{\text{SST}}$, then

$$\begin{aligned}
&C_\alpha(T_{new}) \leq C_\alpha(T_{old}) \\
&\iff C_{imp}(T_{new}) + \alpha M \text{SST} \leq C_{imp}(T_{old}) + \alpha(M+1)\text{SST} \\
&\iff \frac{C_{imp}(T_{new})}{\text{SST}} \leq \frac{C_\alpha(T_{old})}{\text{SST}} + \alpha \\
&\iff R^2_{new} - R^2_{old} \geq \alpha
\end{aligned}$$