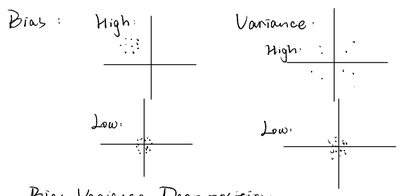
Bias and Variance

1. Generalization Error



Bias-Variance Decomposition

Regression:

$D: \{x_i, y_i\}_{i=1}^{N}$

Determine the weight w.

$\Rightarrow \quad y = w^T \phi(x)$

New Statement:

$S = \{(x_i, y_i)\}_{i=1}^{m}$, where $y_i \in \mathbb{R}$

A hypothesis class $\mathcal{H}$ of regression functions.

e.g. $\mathcal{H} = \{h(x) = w^T\phi(x) \mid w\}$. "假设空间" ("w is arbitrary")

$\Rightarrow$ Choose one hypothesis h from $\mathcal{H}$.

Training Error

$\hat{\varepsilon}(h) = \frac{1}{m}\sum_{i=1}^{m}(y_i - h(x_i))^2$

Training:

Objective: $\hat{h} = \arg\min_{h \in \mathcal{H}} \hat{\varepsilon}(h)$

$\Rightarrow$ training error: $\hat{\varepsilon}(\hat{h})$

D (population) —"i.i.d"→ S_1, S_2, S_3 ... S_K (Training Sets) —"learning algorithm"→ h_{S_1}, h_{S_2}, h_{S_3}, h_{S_K}

Generalization Error:

$\varepsilon(h_S) = E_{(x,y)\sim D}[(y - h(x))^2]$

Expectation (sampling (x,y) from D)

Generalization Gap.

$\varepsilon(h_S) - \hat{\varepsilon}(h_S)$

$\varepsilon = E_S[\varepsilon(h_S)] = E_S\left[E_{(x,y)\sim D}[(y - h(x))^2]\right]$

ε: Expected generalization error $\rightarrow$ We want to minimize it!

☆ The Bias-Variance Decomposition

$\varepsilon = E_S[E_{(x,y)}[(y - h_S(x))^2]]$

1. $h_S(x) \rightarrow h_S$ for readability

$\varepsilon = E_S\left[E_{(x,y)\sim D}[(y - h_S)^2]\right]$

$= E_S E_{(x,y)}[(y - E_S(h_S) + E_S(h_S) - h_S)^2]$

$= E_S E_{(x,y)}[(y - E_S(h_S))^2] + E_S E_{(x,y)}[(E_S(h_S) - h_S)^2]$

$+ 2E_S E_{(x,y)}[\underbrace{(y - E_S(h_S))}_{\text{scalar (for } E_S\text{)}}(E_S(h_S) - h_S)]$

$= E_S E_{(x,y)}[(y - E_S(h_S))^2] + E_S E_{(x,y)}[(E_S(h_S) - h_S)^2]$

$+ 2E_{(x,y)}[(y - E_S(h_S)) \cdot (E_S(E_S(h_S)) - E_S h_S)]$

☆ $E_S(E_S(h_S)) = E_S(h_S)$!

$\Rightarrow E_S(E_S(h_S)) - E_S h_S = 0$!

$= E_S E_{(x,y)}[(y - E_S(h_S))^2] + E_S E_{(x,y)}[(E_S(h_S) - h_S)^2]$

$= E_{(x,y)}[(y - E_S(h_S))^2] + E_x E_S[(E_S(h_S) - h_S)^2]$

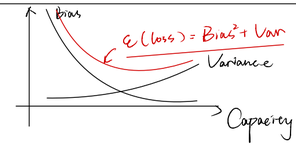
"$(\text{Bias})^2$" "偏离程度" ⟷ "Trade-off" ⟷ "Variance" 离散程度

"Expected Generalization Error = Bias² + Variance!"

illustration:

Impact of Model Capacity on Bias & Variance

☆ Model capacity ↑, Variance ↑ & Bias ↓



Impact of Regularization on Bias & Variance

Since: Regularization ↑ Model Capacity ↓

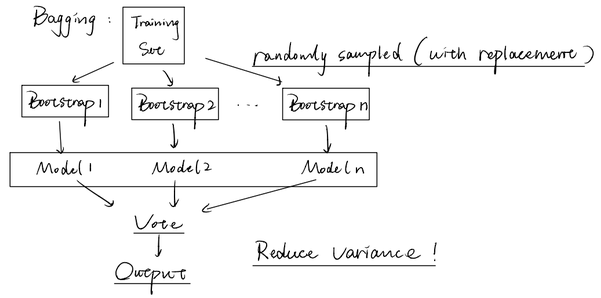
☆ Regularization ↑, Variance ↓ & Bias ↑

Do not over-regularize

Ensemble Learning

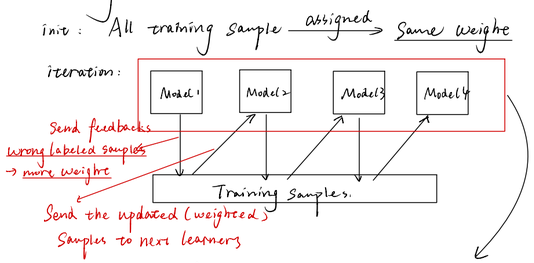
- Bagging: reduce variance
- Boosting: reduce bias.

Bagging:



Reduce variance!

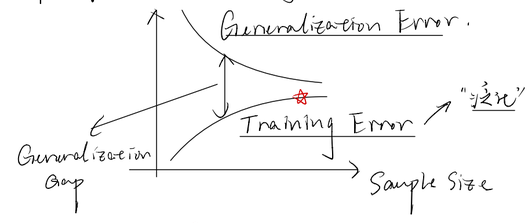
Boosting:



final outcomes: final model = weighted sum of all learners.

"错误率低的模型权重更大"

Impact of Sample Size on Training/Generalization Error:



Sample size ↑ Traning error ↑.

Impact of Sample Size on Bias & Variance:

样本量	对方差(Variance)的影响	对偏差(Bias)的影响	主要风险
非常小	极高 (模型不稳定,随数据波动大)	较低 (取决于模型,但通常能拟合训练点)	过拟合
中等	中等 (开始收敛)	开始反映真实模型能力	取决于模型复杂度
非常大	极低 (模型稳定,泛化能力强)	收敛到固定值 (若模型太简单则偏差高)	欠拟合(如果存在)

Scaling Law:

Performance of LLM increase with:

1. Sample Size
2. Model Capacity
3. GPU Power