$$\underbrace{C_1 \quad C_2 \cdots C_k}_{\text{majority vote}}$$

Ensembler → classifiers

→ improvises your testing error since majority of models arent wrong.

→ Bagging ensembler

↳ Model is just the same eg: Decision tree

Consider a sample set $T_1 \quad T_2 \cdots T_{10000}$

C_1 → sample with replacement $T_1 \quad T_{500} \quad T_1 \quad T_{10} \quad T_{6253} \quad T_{3231} \cdots$

C_2 similarly

⋮

C_k (it's also bagging ensembler)

Or

develop k regression models (varying them)

$M_1 \quad M_2 \cdots M_k$

Errors → $\epsilon_1 \quad \epsilon_2 \cdots \epsilon_k$

$$\frac{\sum \epsilon_i}{k}$$

Let these ϵ_i's follows $N((0,0,0\cdots 0), \operatorname{cov}(\epsilon_1, \epsilon_2, \epsilon_3 \cdots \epsilon_k))$

$$E(\epsilon_1) = 0$$

$$\operatorname{Var}(\epsilon_1) = E(\epsilon_1 - E(\epsilon_1))^2$$

$$= E(\epsilon_1^2) = v$$

An assumption

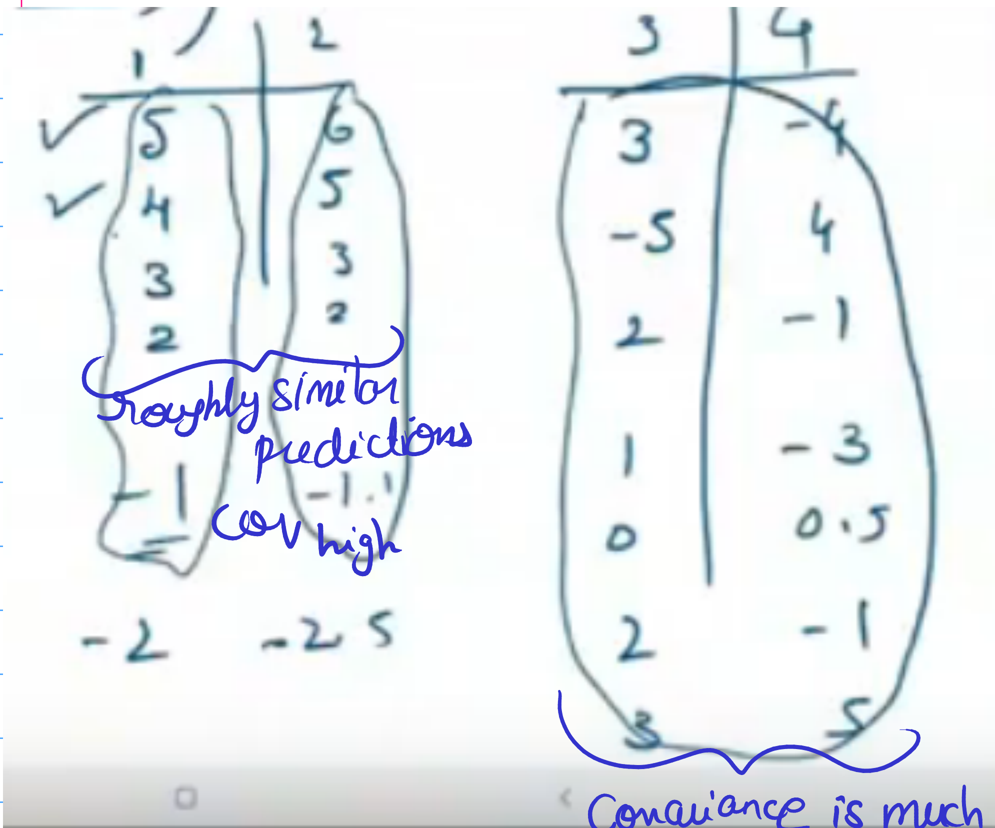
$$Var(\epsilon_1) = Var(\epsilon_2) = \cdots = Var(\epsilon_K)$$
$$= \vartheta$$

$$cov(x,y) = E((x - E(x))(y - E(x)))$$

$\hookrightarrow$ How the variation of x from $E(x)$ is varying w.r.t to variability/variation of y from $E(y)$

$Cov(\epsilon_1, \epsilon_2)$

$\downarrow$



roughly similar predictions

cov high

Covariance is much lesser, they vary independently

$$Cov(\epsilon_1 \epsilon_2) = E(\epsilon_1 - E(\epsilon_1))(\epsilon_2 - E(\epsilon_2))$$
$$= E(\epsilon_1 \epsilon_2)$$
$$= C$$

We assume all covariances are same

$$\begin{array}{c|c}
 & \epsilon_1 \quad \epsilon_2 \quad \cdots \quad \epsilon_k \\
\begin{array}{c}\epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_k\end{array} & \begin{bmatrix} v & c & c & c & c & c \\ c & v & c & c & c & c \\ c & c & c & v & c & \cdots \\ & & & v & & \\ c & c & c & c & v & \\ & & \ddots & & v & \\ & & & & & v \end{bmatrix}
\end{array}$$

$$\begin{array}{cccc} & M_1 & M_2 & M_k \\ \text{Errors} \rightarrow & \epsilon_1 & \epsilon_2 & \epsilon_k \end{array}$$

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_i$$

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_{1i} \quad \cdots \quad \frac{1}{k}\sum_{i=1}^{k} \epsilon_{ni}$$

(1st sample)

Errors: Can you average these?

NO $\quad E\left(\frac{1}{k}\sum_{i=1}^{k} \epsilon_i\right)$

Can cancel

But then isn't error squared already (for argument's sake, lets assume they're not)

$$E\left(\left(\frac{1}{k}\sum_{i=1}^{k} \epsilon_i\right)^2\right)$$

$$= \frac{1}{k^2} E\left(\left(\sum_{i=1}^{k} \epsilon_i\right)^2\right)$$

$$= \frac{1}{k^2} E\left(\sum_{i=1}^{k} \epsilon_i^2 + \sum_i\sum_j \epsilon_i \epsilon_j\right)$$

$$= \frac{1}{k^2} E\left(\sum_{i=1}^{k} \epsilon_i^2\right) + \sum_i\sum_j E(\epsilon_i \epsilon_j)$$

$$= \frac{1}{k^2}\left[E\left(\sum_{i=1}^{k} \epsilon_i^2\right) + \sum_i \sum_j c\right]$$

$$= \frac{1}{k^2}\left[\sum_{i=1}^{k} E(\epsilon_i^2) + \sum_i \sum_j c\right]$$

$$= \frac{1}{k^2}\left[k\,v + k(k-1)\,c\right]$$

$$= \frac{1}{k} v + \frac{k-1}{k} c$$

$\boxed{c = 0}$

if $M_1\ M_2 \cdots M_k$ all behave differently

then $E = \frac{1}{k} v$

M_1 every error variance is v (which is $E(\epsilon_i^2)$)

M_2

$\vdots$

M_2

But the ensemblers together have a variance of $\frac{1}{k}$

But V is a function of ϵ

$\Rightarrow$ even error reduces

$\Rightarrow$ If all models are independent, V comes down by a factor of k

$\Rightarrow$ impressive

if $c \neq 0$

$$c = E(\underbrace{\epsilon_i \epsilon_j}_{\text{if } \epsilon_i \simeq \epsilon_j}) \simeq E(\epsilon_i^2) = v$$

if all ensemblers are same then you get

You common variance to be: $\frac{1}{k}v + \frac{k-1}{k}v$

$$\Rightarrow \quad v \geq \underset{\downarrow}{E(\epsilon_i^2)} \geq \frac{v}{k}$$

$\Rightarrow$ Worst case Ensemblers perform as good as individual models

Similarly $\underline{\underline{C_1}}$ $\underline{\underline{C_2}}$ ---- $\underline{\underline{C_k}}$

$E_{nsembler}(C_1, C_2, \cdots, C_k)$

will perform better, paper mat nikal bc

* It enhances the performance of the model
  - ➡ acts as a "kind of" regularizer without being a regularizer in the strict sense.