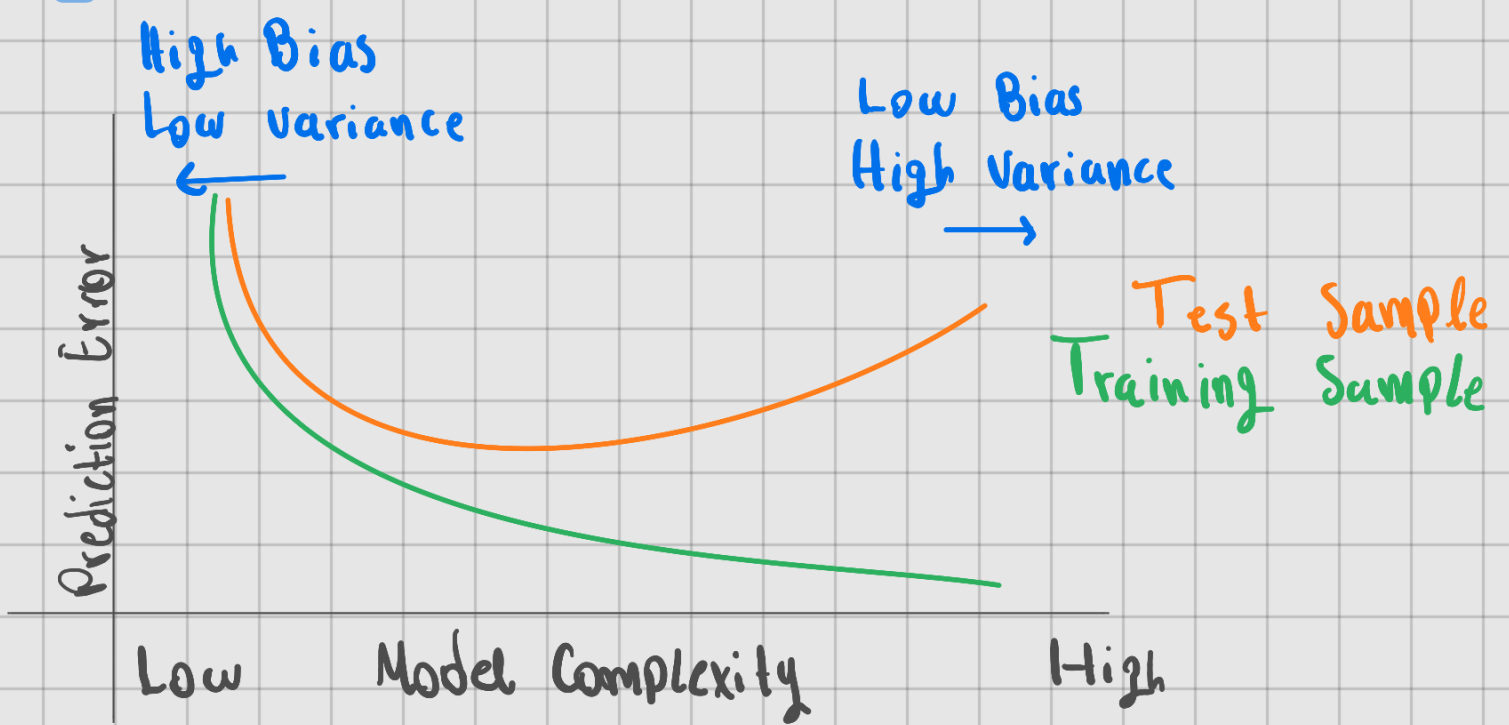


Al momento de entrenamiento tenemos que probar
 Todas las $IG(H, X:t)$



$x^{(1)}$
 $x^{(2)}$
 \vdots
 \vdots
 $x^{(n)}$

80% Entrenamiento

← Datos

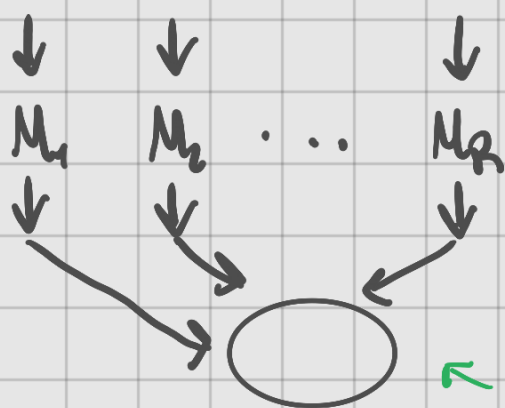
20% Validación

Reducción de Varianza

- Promediar reduce las varianzas

$$\text{Var}(\bar{x}) = \frac{\text{Var}(x)}{N}$$

Combinación de muchos identificadores simples



Bagging

$M_i := \text{Modelo simple}$

Entre todos deciden el mejor identificador

$\Rightarrow M_1 \rightarrow M_2 \rightarrow \dots \rightarrow M_p \rightarrow$

Boosting

El resultado de un modelo se pasa al siguiente

- Se crean varias instancias de los datos de entrenamiento
- Un clasificador por cada sample

Random Forest

Se escoge aleatoriamente las variables (típicamente \sqrt{N} variables) para que todos vean cosas diferentes

Generamos T conjuntos de datos aleatorios con repetición

Tomamos decisiones sacando promedios o por votación

$D \rightarrow \text{Randomize}$



Combinar árboles
 \downarrow
 T^*

