

Ensembles (Ensemble)

X_i $V(x_i) = \sigma^2$

$0 \leq i \leq n$ iid
 \downarrow $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

x_0, x_1, \dots, x_n
 \bar{X}

covar

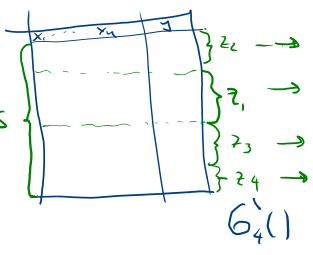
Si no son iid:
$$\text{Var}(\bar{X}) = \rho\sigma^2 + \frac{1-\rho}{n}\sigma^2$$

• Bagging (Conjuntos de datos)

Población: $P \rightarrow$

training set: $S \sim P \rightarrow S = P$

bootstrap set: $Z \sim S$



$z_1, z_2, z_3, \dots, z_m$

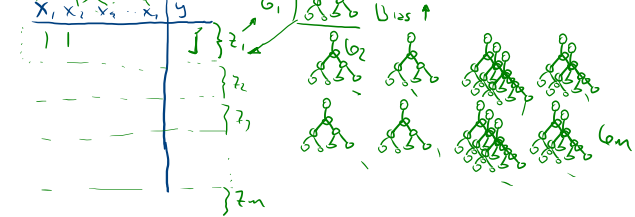
Aggregation:

Modelos de pred: $G_m(z_m)$

Ensemble: $G(x) = \sum_m \frac{G_m(x)}{M}$

$$\text{Var}(\bar{X}) = \rho\sigma^2 + \frac{1-\rho}{M}\sigma^2$$

Random Forests



$$G(x) = \sum_m \frac{G_m(x)}{M}$$

Bagging:

Ventajas:

- menor Varianza (Overfitting menor) ✓

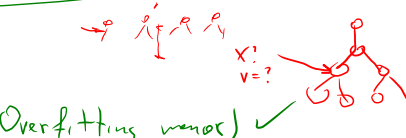
- mayor precisión

Desventajas

- Costoso x

- Interpretabilidad x

- mayor bias x



$z_1 =$ underfitting

$$G(x) = \sum_m \frac{G_m(x)}{M}$$

