Nomenclatura

$\{x_1, x_2, \ldots, x_N\} \equiv$ vectores/amostras

$\{y_1, y_2, \ldots, y_N\} \equiv$ categorias correspondentes

$X_{\mathrm{i}} = (x_i, y_i) \equiv$ padrão de treino

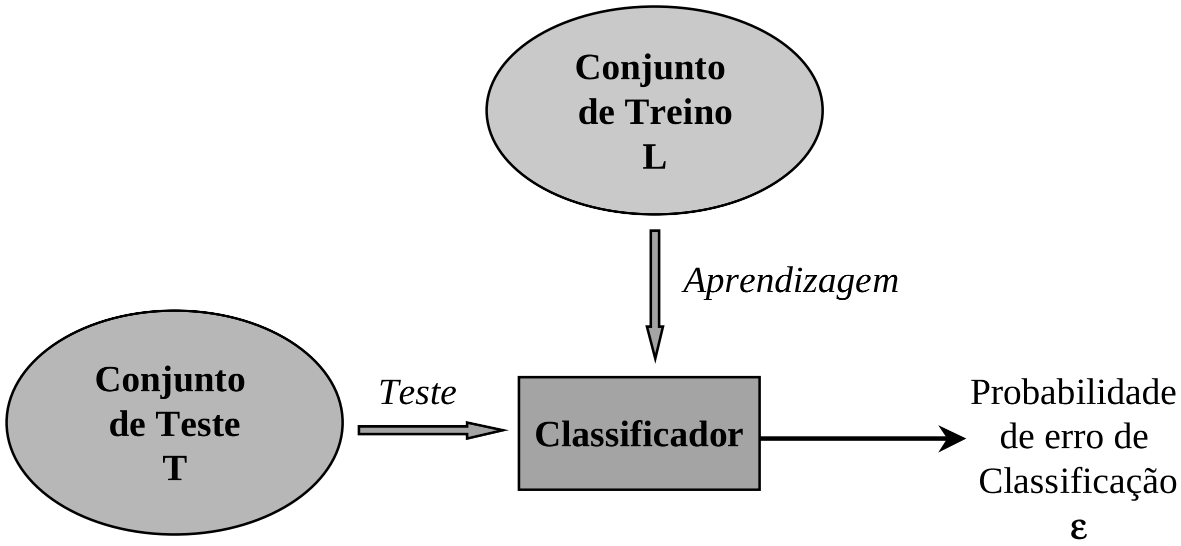
$X = \{X_{\mathrm{i}}, i = 1, \ldots, N\} \equiv$ conjunto de treino

$X_T = \{X_{\mathrm{iT}}, i = 1, \ldots, N_T\} \equiv$ conjunto de teste

$\eta(x, X) \equiv$ regra de decisão

$$n_i = Q[y, \eta(x, X)] = \begin{cases} 0 & \text{se } \eta(x, X) = y \\ 1 & \text{caso contrário} \end{cases} \quad \equiv \text{ indicador de erro}$$

Estimação da Probabilidade de Erro com base em Conjunto Independente de Teste



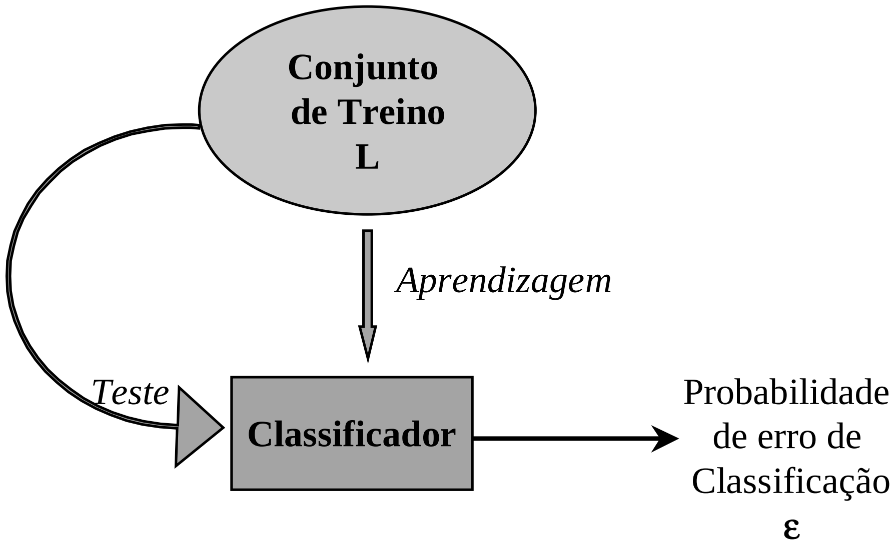
•Contar os erros:

$$n_i = \begin{cases} 0 & \text{se } x_i \text{ é classificado correctamente} \\ 1 & \text{se } x_i \text{ é classificado incorrectamente} \end{cases}$$

$$\boxed{\hat{\varepsilon} = \hat{\varepsilon}(T,L) = \frac{\sum_{i=1}^{N_T} n_i}{N_T} = \frac{\sum_{i=1}^{N_T} Q\left[y_{iT}, \eta(x_{iT}, X)\right]}{N_T}}$$

$$\hat{\varepsilon}(T,L) > \varepsilon$$

Erro Aparente ou de Re-substituição



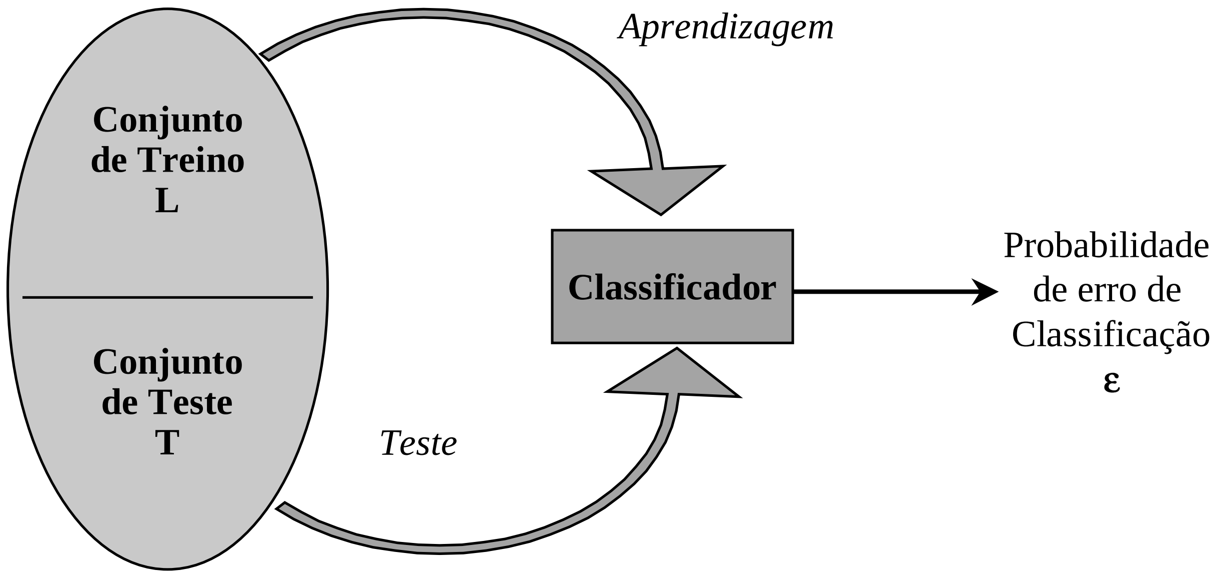
$$\hat{\varepsilon}_{app} = \frac{\sum_{i=1}^{N} Q[y_i, \eta(x_i, X)]}{N}$$

- O mesmo conjunto de treino é usado para avaliação do classificador
- As estimativas da probabilidade de erro assim calculadas são optimistas:

$$\hat{\varepsilon}_{app} = \hat{\varepsilon}(L, L) < \varepsilon$$

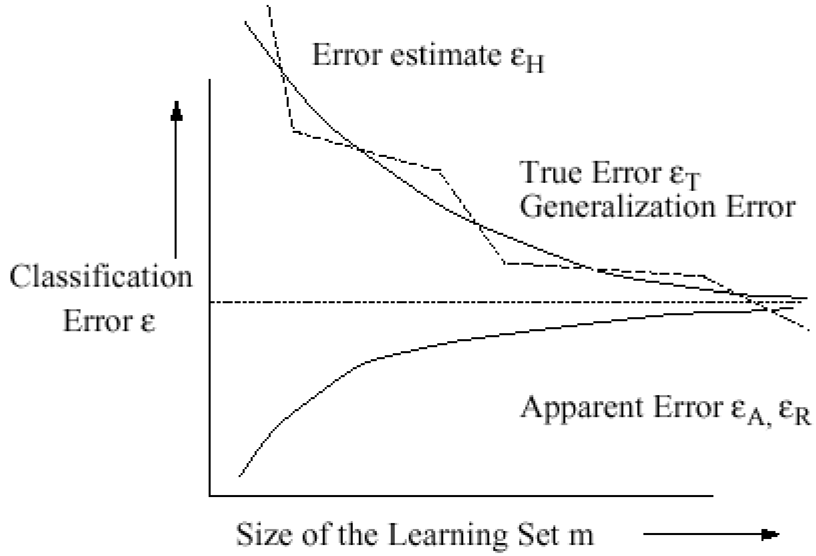
Este efeito é tanto mais notório quanto menor for o conjunto de treino *(over-fitting)*

Método *Hold-Out*

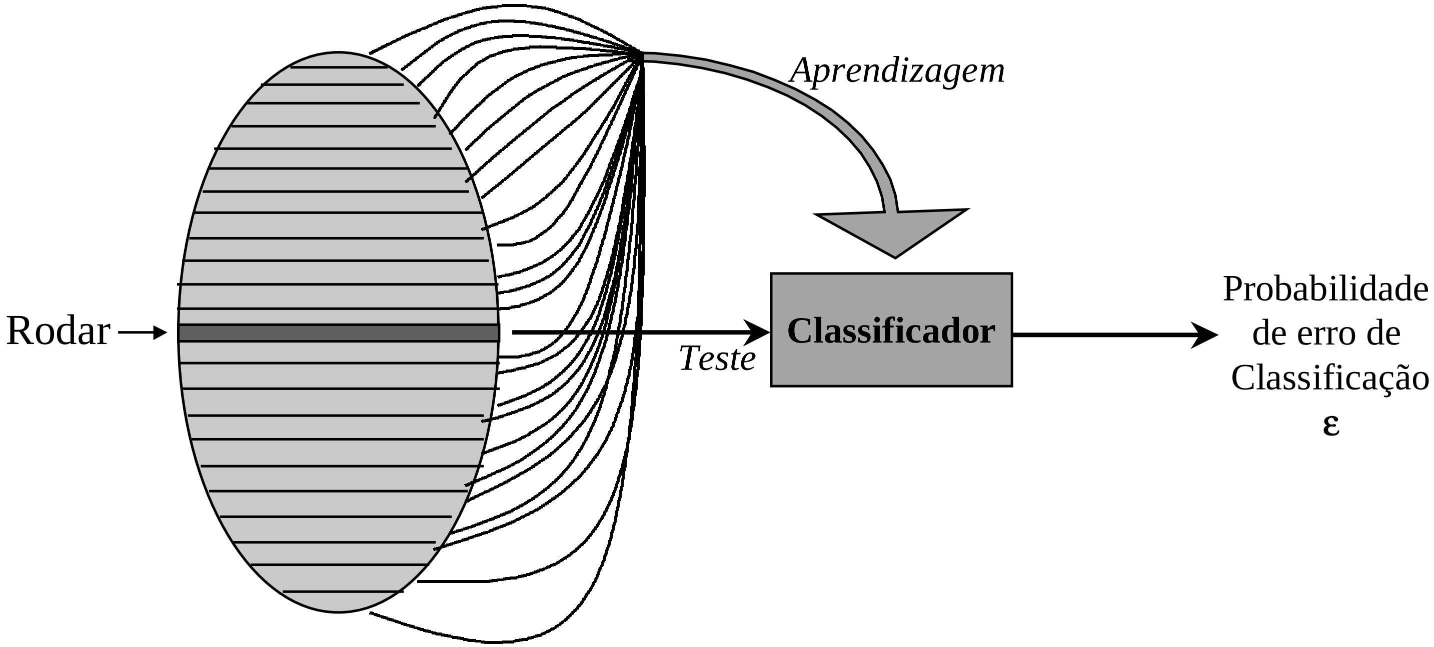


$$\hat{\varepsilon}_H = \hat{\varepsilon}(T, L) = \frac{\sum_{i=1}^{N_T} Q\left[y_{iT}, \eta(x_{iT}, X_L)\right]}{N_T}$$

$\hat{\varepsilon}_H(T, L) > \varepsilon$: pessimista.


Error estimate ε_H
True Error ε_T
Generalization Error
Classification
Error ε
Apparent Error $\varepsilon_{A,}$ ε_R
Size of the Learning Set m

Método *Leave-one-Out* ou *Jack-knifing* ou *Cross-Validation*



- Guarda uma amostra para teste, usando os restantes N-1 padrões para treino

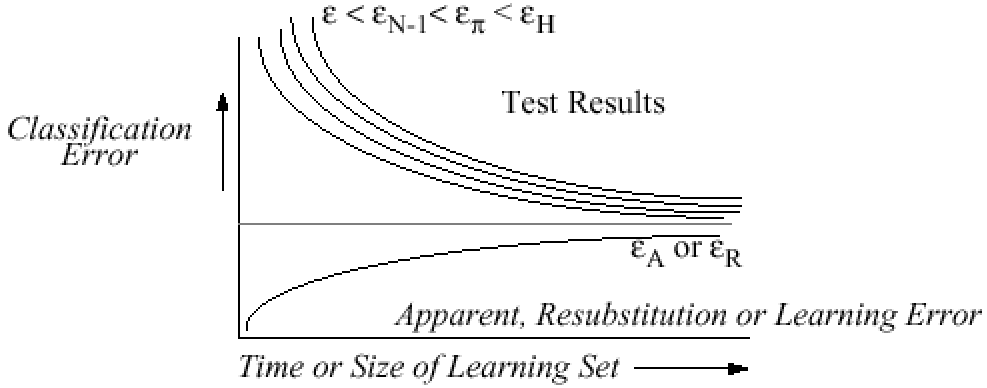
$$\hat{\varepsilon}_{JK} = \frac{\sum_{i=1}^{N} Q\left[y_i, \eta(x_i, X_{(i)})\right]}{N}$$

$$c/ \quad X_{(i)} = X - \{X_i\} \text{ (retirar o } j\text{ésimo padrão de } X\text{)}$$

  – Estimativa menos polarizada que a anterior mas de grande variância

Rotação e Estimação pelo Método *Leave-one-out*

- **Particionar o conjunto de dados num número elevado de sub-conjuntos (no limite, N)**
- **Usar todos os sub-conjuntos à excepção de um para treino**
- **Usar o conjunto deixado de fora para teste**
- **Repetir sobre todos os sub-conjuntos calculando a média das estimativas**


ε < ε_{N-1} < ε_π < ε_H
Test Results
Classification Error
ε_A or ε_R
Apparent, Resubstitution or Learning Error
Time or Size of Learning Set

Bootstraping

Mundo Real

Modelo probabilístico desconhecido

Dados observados

$$P \longrightarrow X = \{x_1, x_2, \ldots, x_N\}$$

$$\theta = \theta(P) \qquad \hat{\theta} = S(X)$$

Parâmetros de interesse

Estimativa

$\Longrightarrow$

Mundo Bootstrap

Estimativa do modelo probabilístico

Conjunto de bootstrap

$$\hat{P} \longrightarrow X^* = \{x_1^*, x_2^*, \ldots, x_N^*\}$$

$$\hat{\theta} = \theta(\hat{P}) \qquad \hat{\theta}^* = S(X^*)$$

Parâmetros de interesse

Réplica de $\hat{\theta}$ por bootstrap

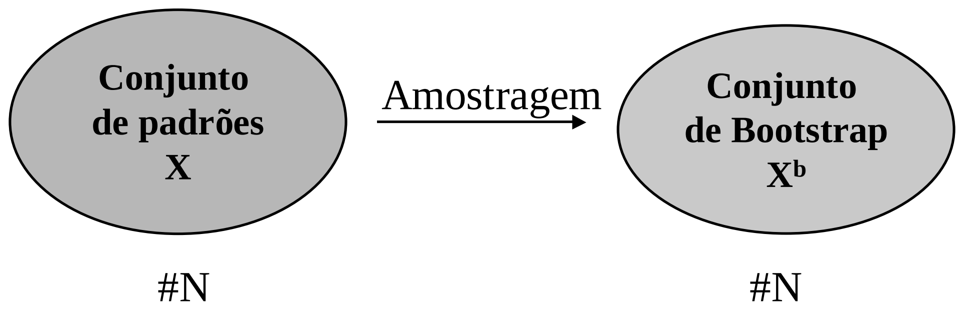
Técnicas de *Bootstraping* na Estimação da Probabilidade de Erro do Classificador

- **Uso de conjuntos de treino obtidos por amostragem, com reposição, dos dados/padrões existentes**

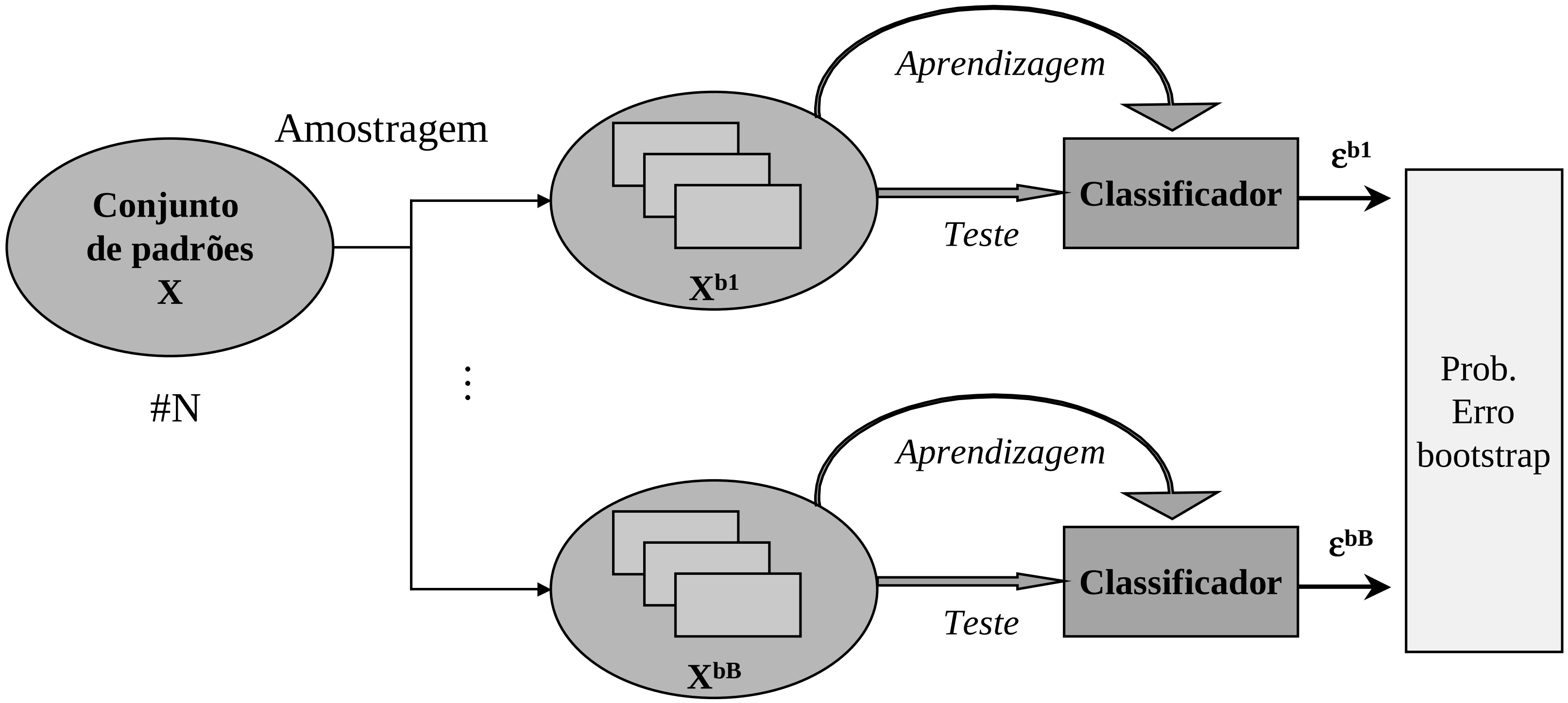
$$X = \{X_i, i = 1, \dots, N\} \qquad X^b = \left\{ \underbrace{X_3, X_1, X_5, \dots, X_1, \dots}_{\#N} \right\}$$

Conj. de *bootstrap*;

Conj. de *N* padrões seleccionados aleatoriamente de X

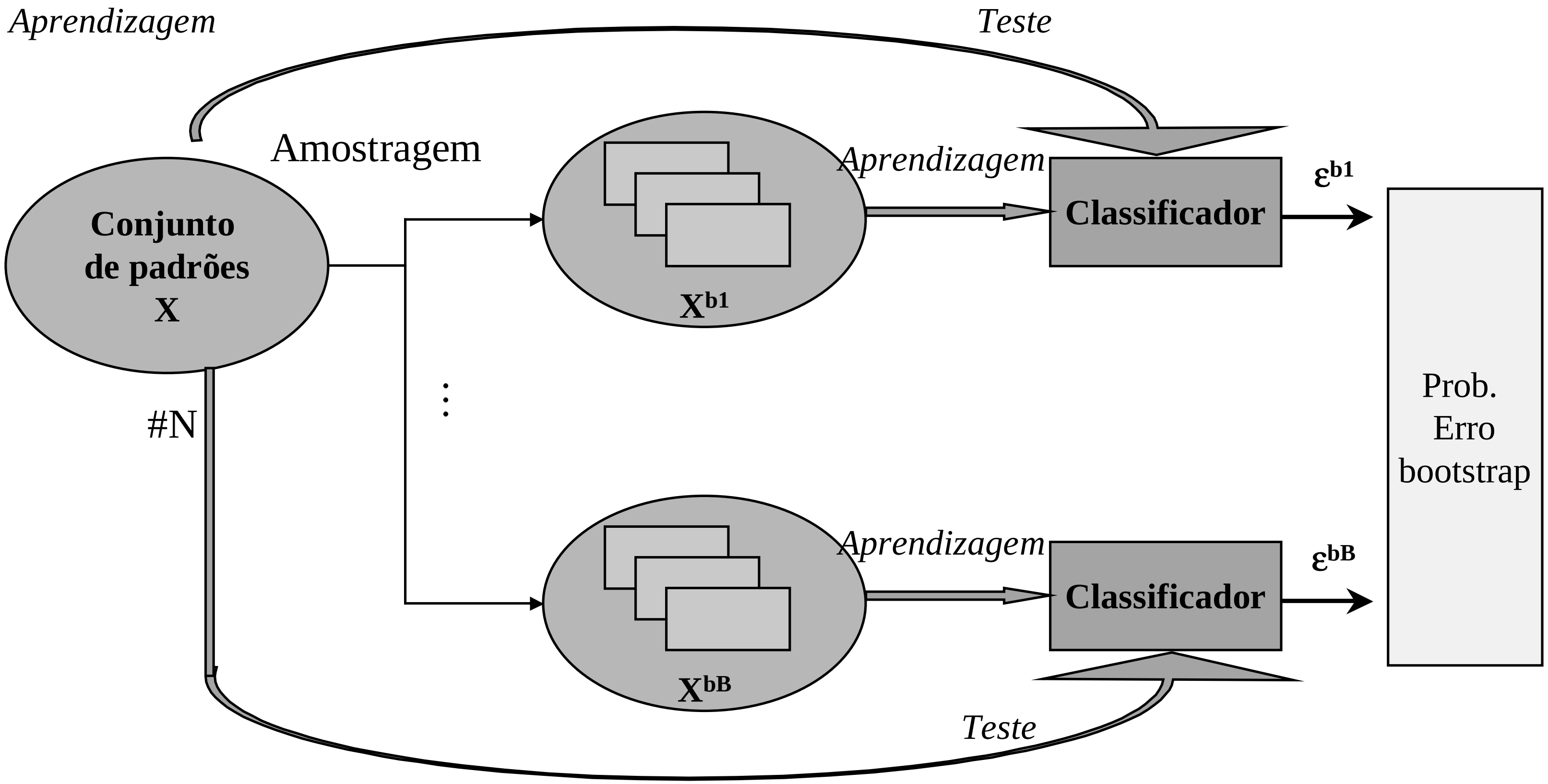


Erro Aparente por Bootstrap



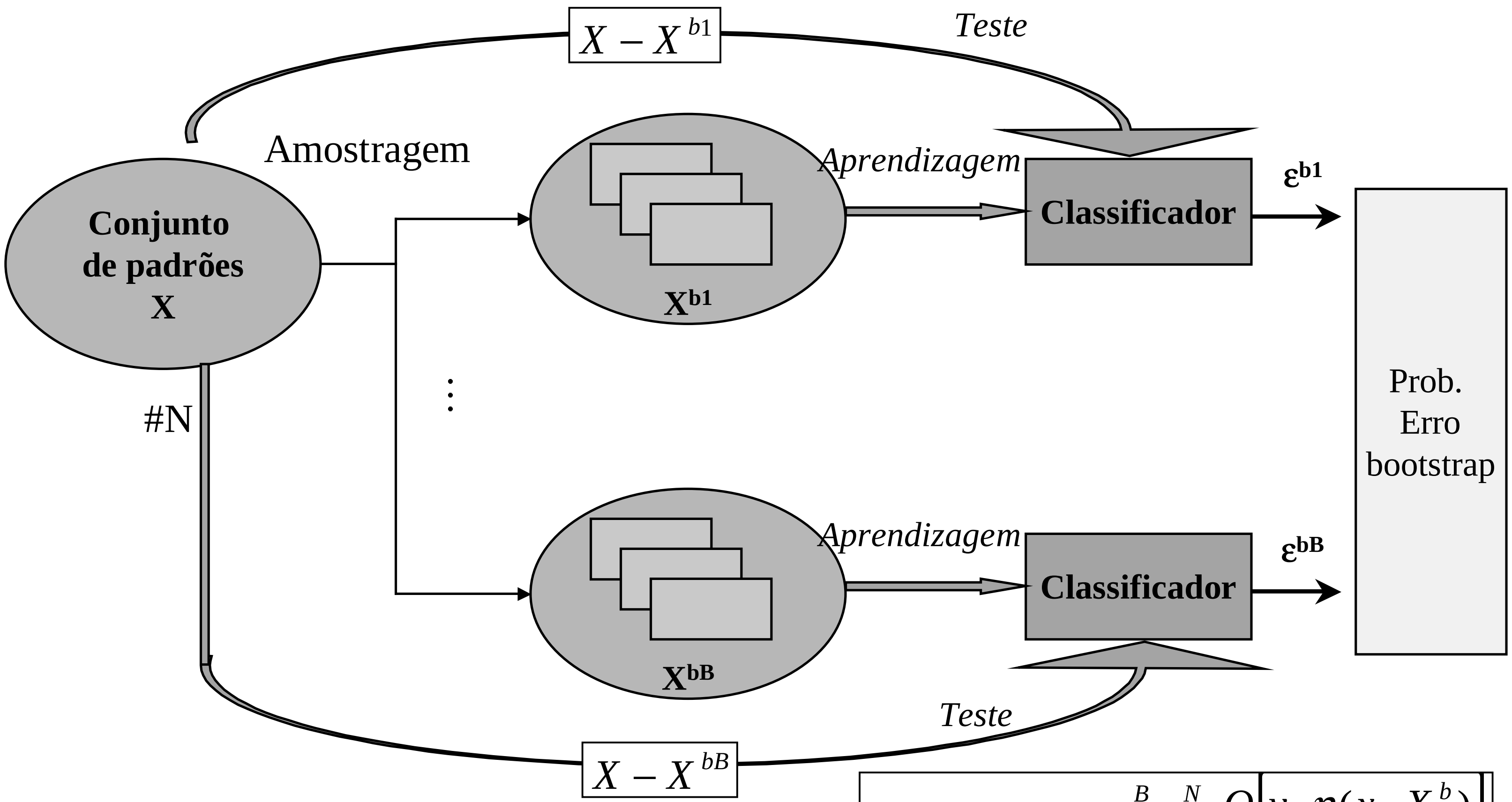
$$\hat{\varepsilon}_{Boot-app} = App = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N}\sum_{\substack{i=1 \\ x_i \in X^b}}^{N} Q\left[y_i, \eta(x_i, X^b)\right]$$

Erro Simples Estimado por *Bootstrap*



$$\hat{\varepsilon}_{Boot-simple} = Simple = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N}\sum_{\substack{i=1 \\ x_i \in X}}^{N} Q\left[y_i, \eta(x_i, X^b)\right]$$

Estimativa *Jack-Knifing* por *Bootstrap*



$$\hat{\varepsilon}_{Boot-JK} = E_o = \sum_{b=1}^{B} \sum_{A_b}^{N} \frac{Q\left[y_i, \eta(x_i, X^b)\right]}{\sum_{b=1}^{B} |A_b|}$$

$$c/ \quad A_b = \left\{ i \mid X_i \in X \wedge x_i \notin X^b \right\}$$

Estimativa *E632*

$$\boxed{E_{632} = 0.386 \times A_{pp} + 0.362 \times E_o}$$

- **Ideia: ajuste no optimismo do erro aparente**
- **Factor 0.362 provêm de um argumento teórico que as amostras de bootstrap usadas no cálculo de Eo estão distânciadas da média de um conjunto de teste típico por um factor aproximadamente de 1/.362**
- **Desta forma a estimativa torna-se praticamente não enviezada**