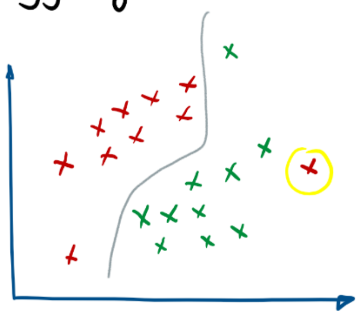
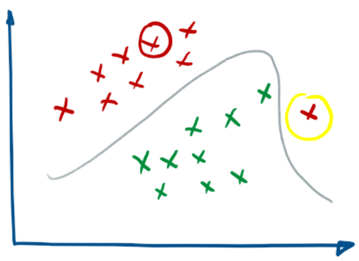
Notes of Ensemble Learning

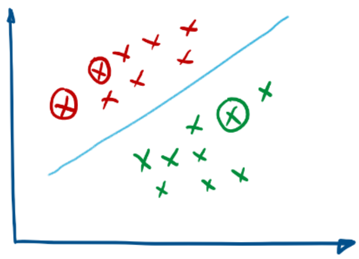
Zengchang Qin (Ph.D.)

1) Bagging

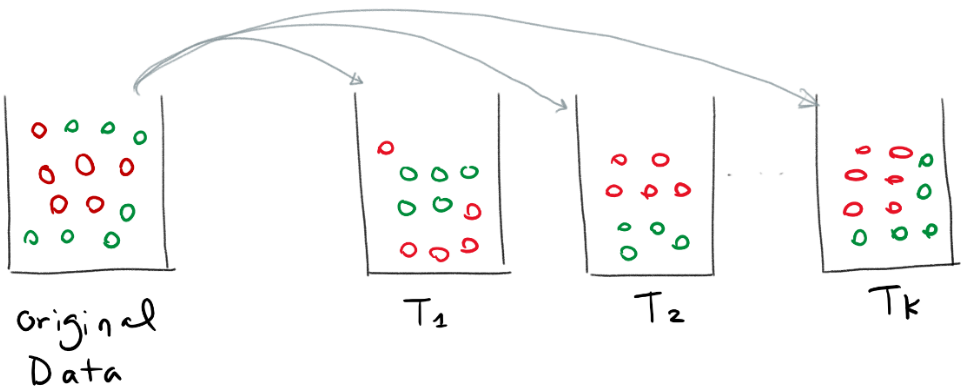


Sampling with replacement, noise has a lower probability to be sampled!





2) Random Forest : Bagging with CARTs.



$$F(x^{(i)}) = \frac{1}{K} T_k(x^{(i)})$$

For a general case :

$$F(x^{(i)}) = \frac{1}{K} M_k(x^{(i)})$$

where M_k could be any classification model.

3) Boosting

Given the training data $D = \{(x^{(1)}, y^{(1)}), \cdots (x^{(N)}, y^{(N)})\}$

$x^{(i)} \in R^n, \quad y \in \{+1, -1\}$

For each data $x^{(i)}$, there is an associated weight $w^{(i)} \in R$

3.1 Initialization

$$W = (w^{(1)}, w^{(2)}, \cdots w^{(N)})$$
$$w^{(i)} = \frac{1}{N}, \quad i = 1, 2 \cdots N$$

3.2 For $k = 1, 2, \cdots K$

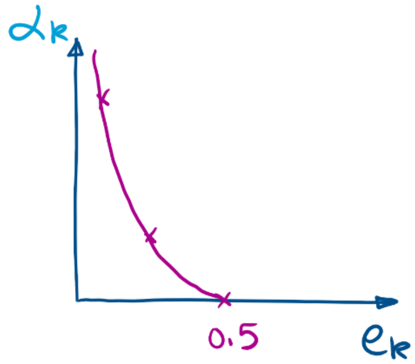
$G_k(x)$ is a weak classifier (E.g. N.B. or DT)

The error for each $G_k(x)$

$$e_k = P(G_k(x^{(i)}) \neq y^{(i)})$$
$$= \sum_i w^{(i)} I(G_k(x^{(i)}) \neq y^{(i)})$$

The weight for classifier $G_k(x)$ is updateb by:

$$\alpha_k = \frac{1}{2} \ln\left(\frac{1 - e_k}{e_k}\right)$$



$e_k = 0.5$ (Random classifier)

$$\alpha_k = \frac{1}{2} \ln\left(\frac{0.5}{0.5}\right) = 0$$

$e_k = 0.2$ (Good classifier)

$$\alpha_k = \frac{1}{2} \ln\left(\frac{0.8}{0.2}\right) = \frac{1}{2} \ln 4$$

$e_k = 0.1$ (Excellent classifier)

$$\alpha_k = \frac{1}{2} \ln\left(\frac{0.9}{0.1}\right) = \frac{1}{2} \ln 9$$

3.3 Updating W_k, the weights of data at round k.

$$W_k = (w_k^{(1)}, w_k^{(2)} \cdots w_k^{(N)})$$

$$w_{k+1}^{(i)} = \frac{w_k^{(i)}}{Z_k} \exp(-\alpha_k y^{(i)} G_k(x^{(i)}))$$

if $y^{(i)} = G_k(x^{(i)})$ $\quad w_k^{(i)} \cdot \frac{1}{e^{\alpha_k}}$ ↓

if $y^{(i)} \neq G_k(x^{(i)})$ $\quad w_k^{(i)} \cdot e^{\alpha_k}$ ↑

we tend to enlarge the weights for misclassified data. Decrease the weights of data with correct classification.

Z_k is the normalization factor:

$$Z_k = \sum_{i=1}^{N} w_k^{(i)} \exp(-\alpha_k G_k(x^{(i)}))$$

Because of Z_k, W_k is a distribution of importance of data, for $k = 1 \cdots K$,

3.4 $f(x^{(i)}) = \sum_{k=1}^{K} \alpha_k G_k(x^{(i)})$

$G(x^{(i)}) = \text{sign}(f(x^{(i)}))$ in final classification.