Boosting and Gradient Boosted Trees

  Romeo Kienzler

Boosting

Boosting

aka. “Stagewise Additive Modeling”

Boosting

aka. “Stagewise Additive Modeling”

$$h_i(x)$$

Boosting

aka. “Stagewise Additive Modeling”

$$h_i(x)$$

Boosting

aka. “Stagewise Additive Modeling”

$$h_i(x)$$

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x)$$

Boosting

aka. “Stagewise Additive Modeling”

$$h_i(x)$$

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x)$$

Boosting

aka. "Stagewise Additive Modeling"

$$h_i(x)$$

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x)$$

Boosting

aka. “Stagewise Additive Modeling”

$$h_i(x)$$

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x)$$

Boosting

aka. "Stagewise Additive Modeling"

$$h_i(x)$$

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Boosting

aka. "Stagewise Additive Modeling"

$$h_i(x)$$

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Romeo Kienzler

Boosting

aka. “Stagewise Additive Modeling”

$$h_i(x)$$

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Boosting

aka. "Stagewise Additive Modeling"

$$h_i(x)$$

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Boosting

aka. “Stagewise Additive Modeling”

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. "Stagewise Additive Modeling"

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. "Stagewise Additive Modeling"

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. "Stagewise Additive Modeling"

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. “Stagewise Additive Modeling”

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. "Stagewise Additive Modeling"

$$L(w) = \frac{1}{2N}\sum_{n=1}^{N}(y_n - (Xw)_n)^2$$

Romeo Kienzler

Boosting

aka. "Stagewise Additive Modeling"

$$L(w) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (Xw)_n)^2$$

Boosting

aka. "Stagewise Additive Modeling"

$$L(w) = \frac{1}{2N}\sum_{n=1}^{N}(y_n - (Xw)_n)^2$$

Romeo Kienzler

Boosting

aka. “Stagewise Additive Modeling”

$$L(w) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (Xw)_n)^2$$

  Romeo Kienzler

Boosting

aka. “Stagewise Additive Modeling”

$$L(w) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (Xw)_n)^2$$

Boosting

aka. “Stagewise Additive Modeling”

$$\frac{\partial L}{\partial w_j} = -\frac{1}{N}\sum_{n=1}^{N} x_{nj}(y_n - (Xw)_n)$$

$$L(w) = \frac{1}{2N}\sum_{n=1}^{N}(y_n - (Xw)_n)^2$$

  Romeo Kienzler

Boosting

aka. “Stagewise Additive Modeling”

$$\frac{\partial L}{\partial w_j} = -\frac{1}{N}\sum_{n=1}^{N} x_{nj}(y_n - (Xw)_n)$$

$$L(w) = \frac{1}{2N}\sum_{n=1}^{N}(y_n - (Xw)_n)^2$$

Boosting

aka. "Stagewise Additive Modeling"

$$f(x) = g(x) + h(x) \quad \rightarrow \quad f'(x) = g'(x) + h'(x)$$

$$L(w) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (Xw)_n)^2$$

Boosting

aka. “Stagewise Additive Modeling”

$$\frac{\partial L}{\partial w}$$

$$f(x) = x^n \quad \rightarrow \quad f'(x) = n \cdot x^{n-1}$$

$$L(w) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (Xw)_n)^2$$

Boosting

aka. "Stagewise Additive Modeling"

$$\frac{\partial L}{\partial w_j} = -\frac{1}{N}\sum_{n=1}^{N} x_{nj}(y_n - (Xw)_n)$$

$$L(w) = \frac{1}{2N}\sum_{n=1}^{N}(y_n - (Xw)_n)^2$$

Boosting

aka. “Stagewise Additive Modeling”

$$f(x) = x \quad \rightarrow \quad f'(x) = 1$$

$$\frac{\partial L}{\partial w_j} = -\frac{1}{N} \sum x_{nj}(y_n - (Xw)_n)$$

$$f(x) = g(x) \cdot h(x) \quad \rightarrow \quad f'(x) = g'(x) \cdot h(x) + g(x) \cdot h'(x)$$

$$L(w) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (Xw)_n)^2$$

Boosting

aka. “Stagewise Additive Modeling”

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. "Stagewise Additive Modeling"

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. “Stagewise Additive Modeling”

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Boosting

aka. “Stagewise Additive Modeling”

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

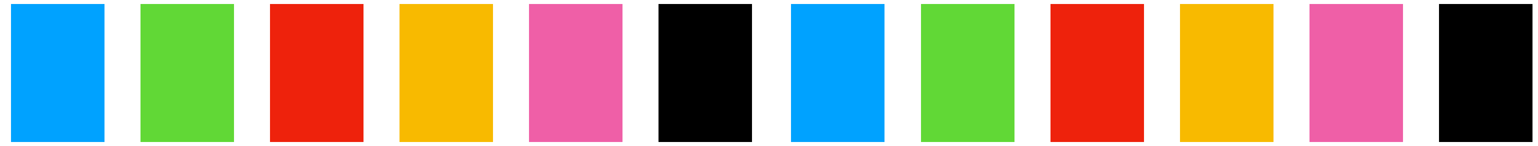
$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

Gradient Boosted Trees

$$F_0(X) = h_0 = mean(y)$$

$$F_0(X) = h_0 = mean(y)$$

$$F_0(X) = h_0 = mean(y)$$



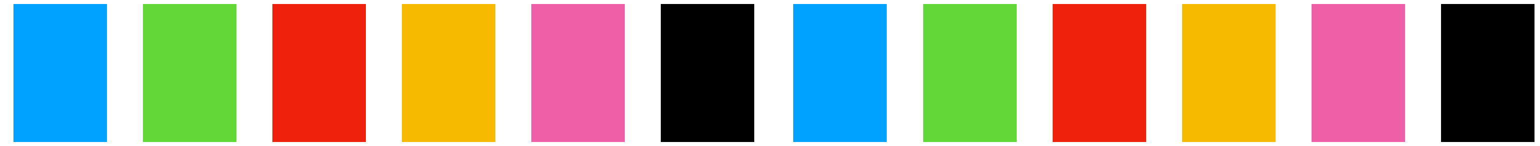
Romeo Kienzler

$$F_0(X) = h_0 = mean(y)$$

$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$F_0(X) = h_0 = mean(y)$

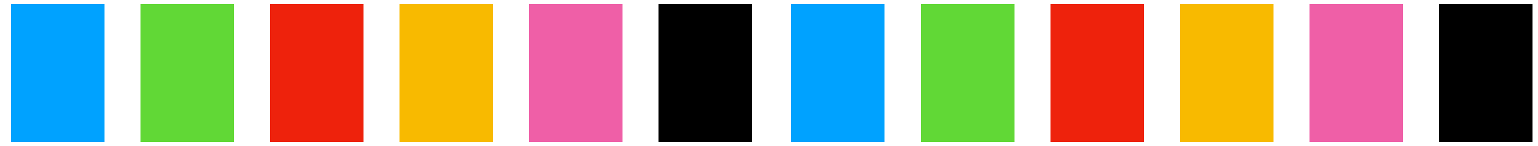


$$\epsilon_1 = y - F_0(X)$$

$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_0(X) = h_0 = mean(y)$$

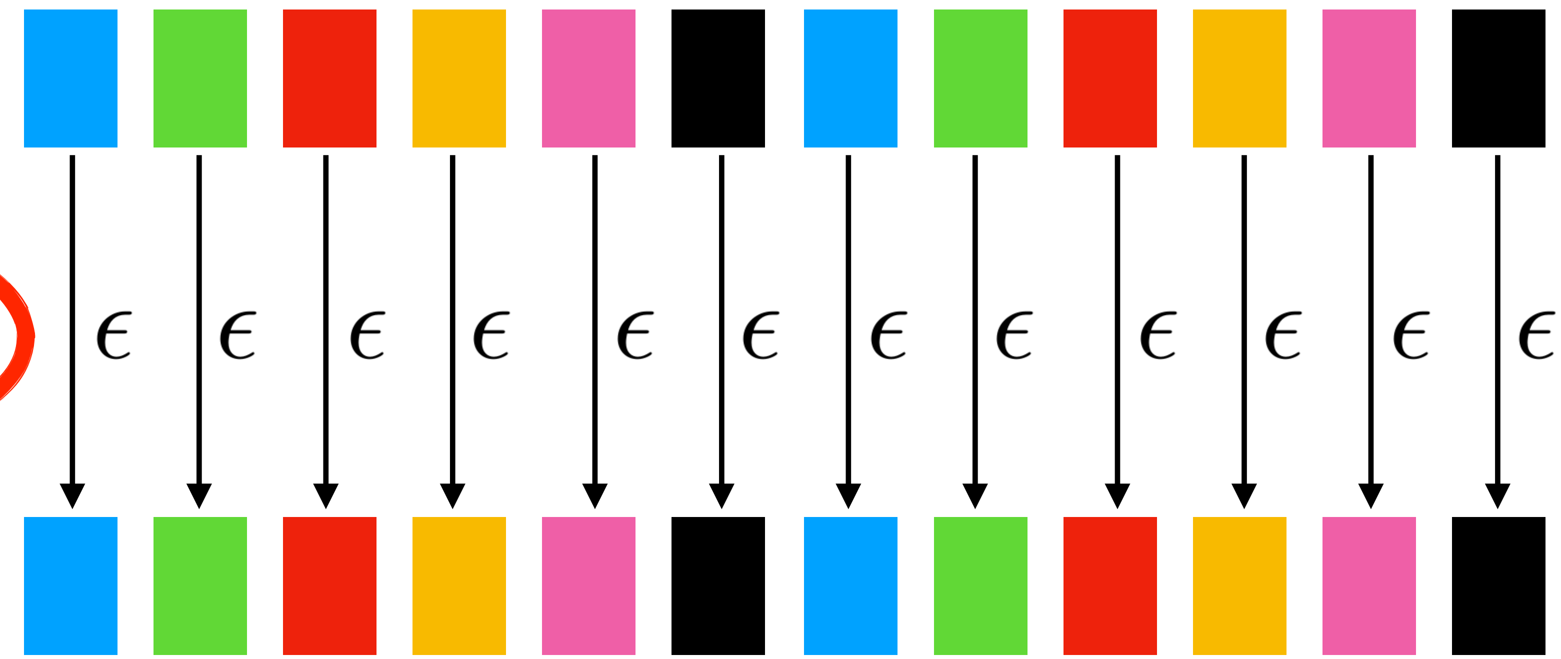


$$\epsilon_1 = y - F_0(X)$$

Romeo Kienzler

$$F_0(X) = h_0 = mean(y)$$

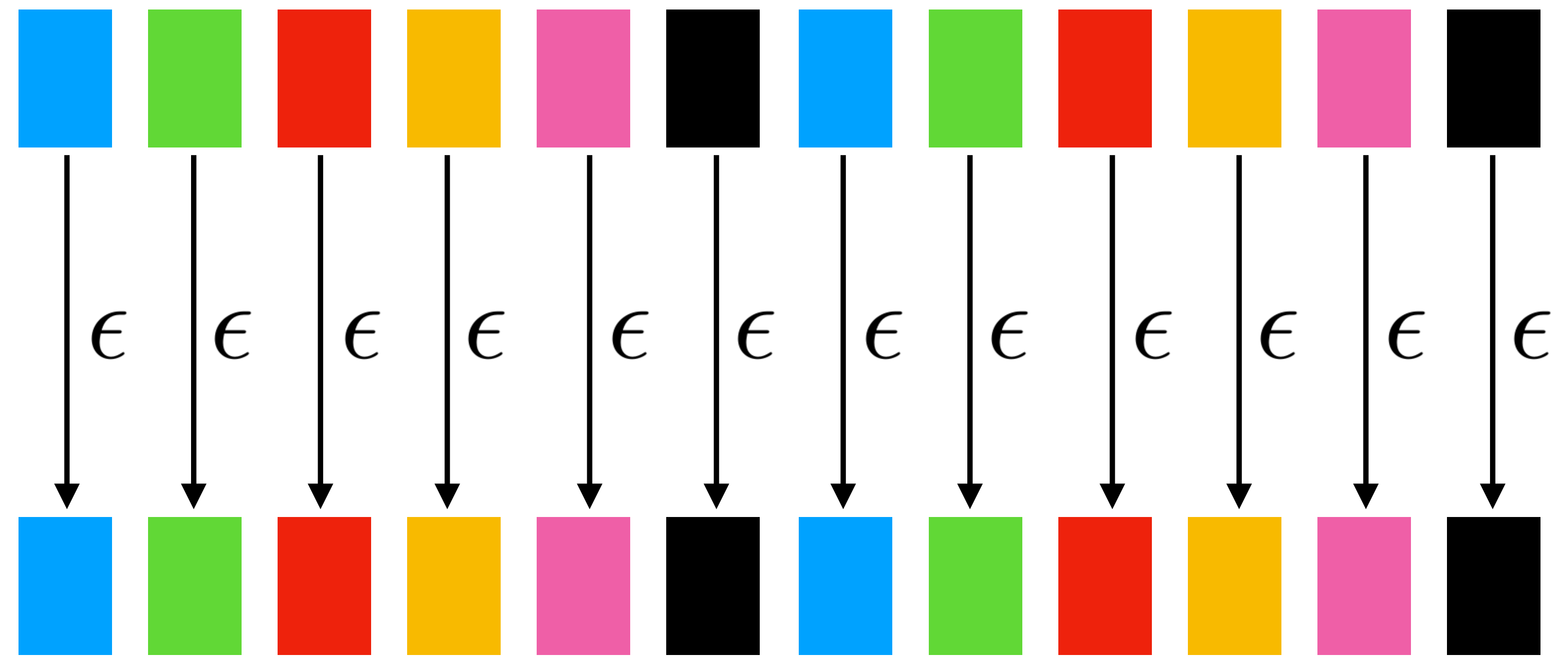
$$\epsilon_1 = y - F_0(X)$$



$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

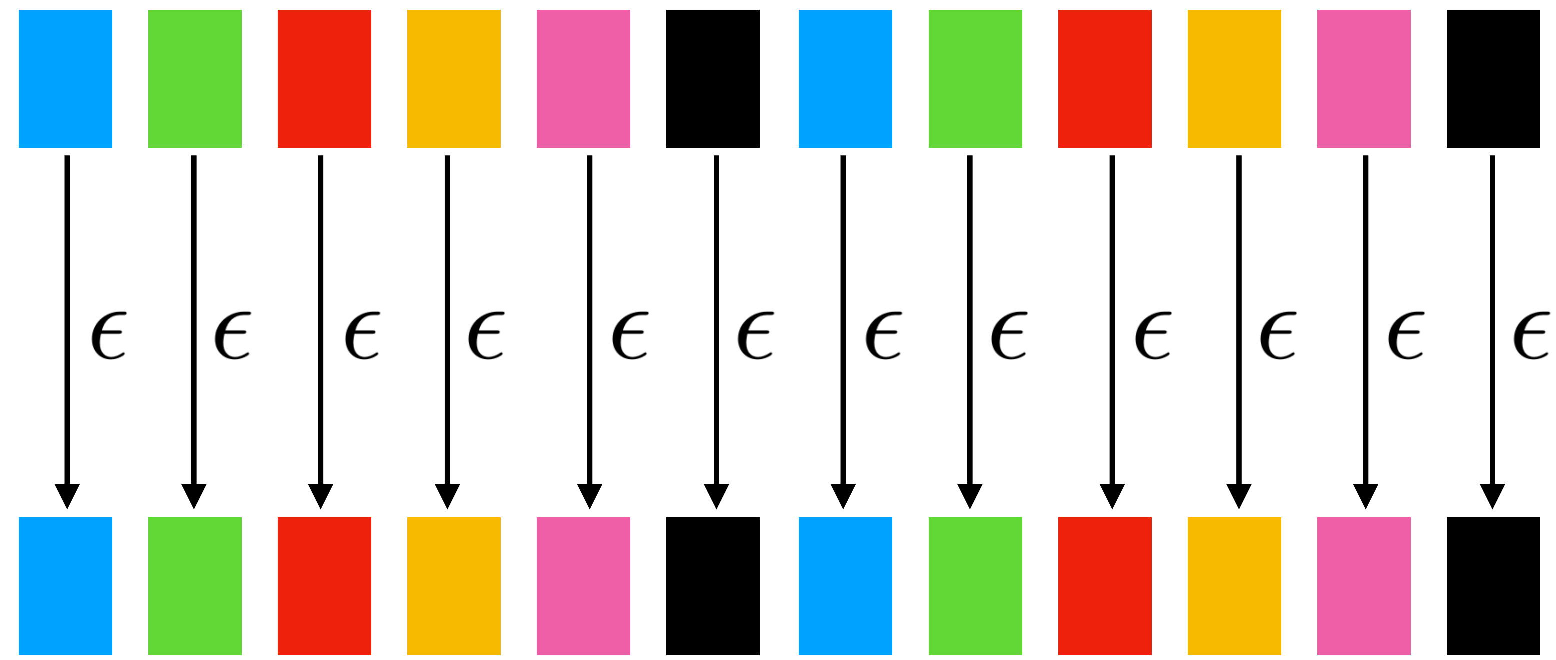
$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$



$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

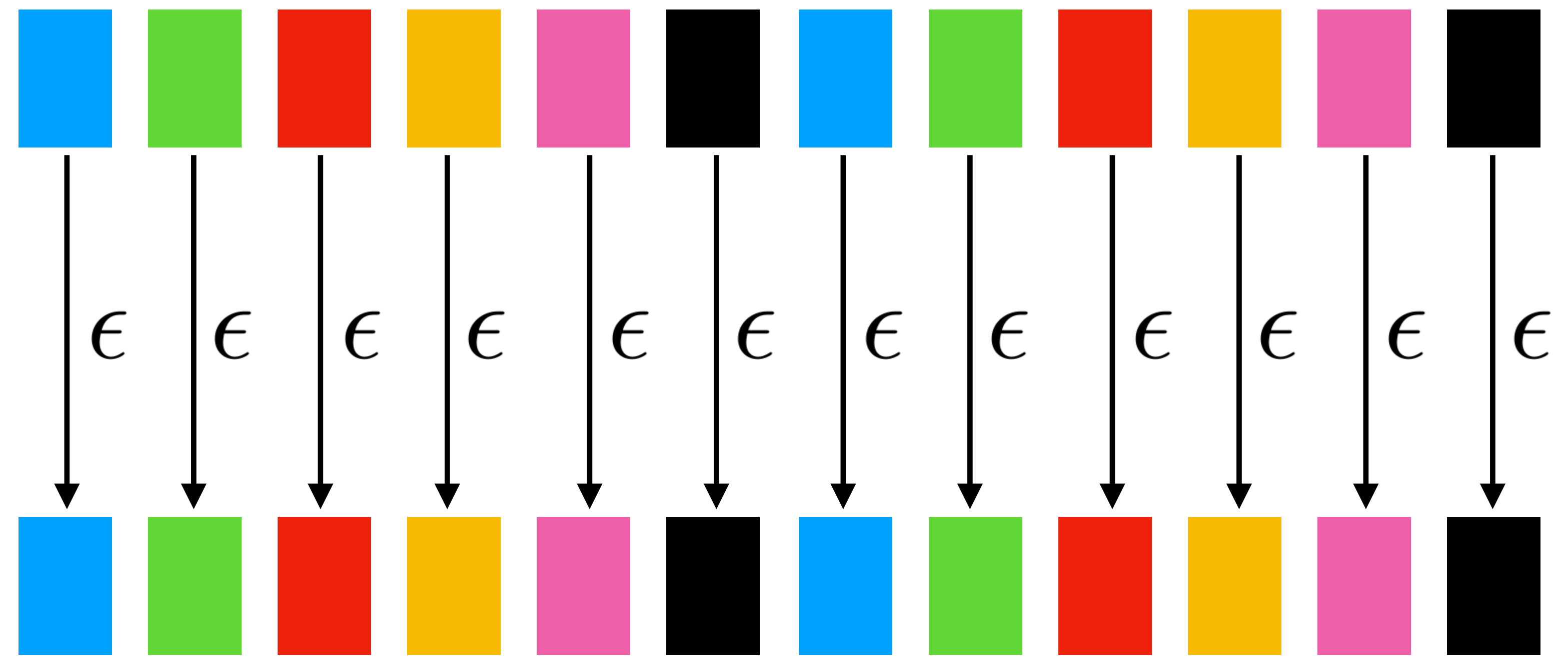


  Romeo Kienzler

$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

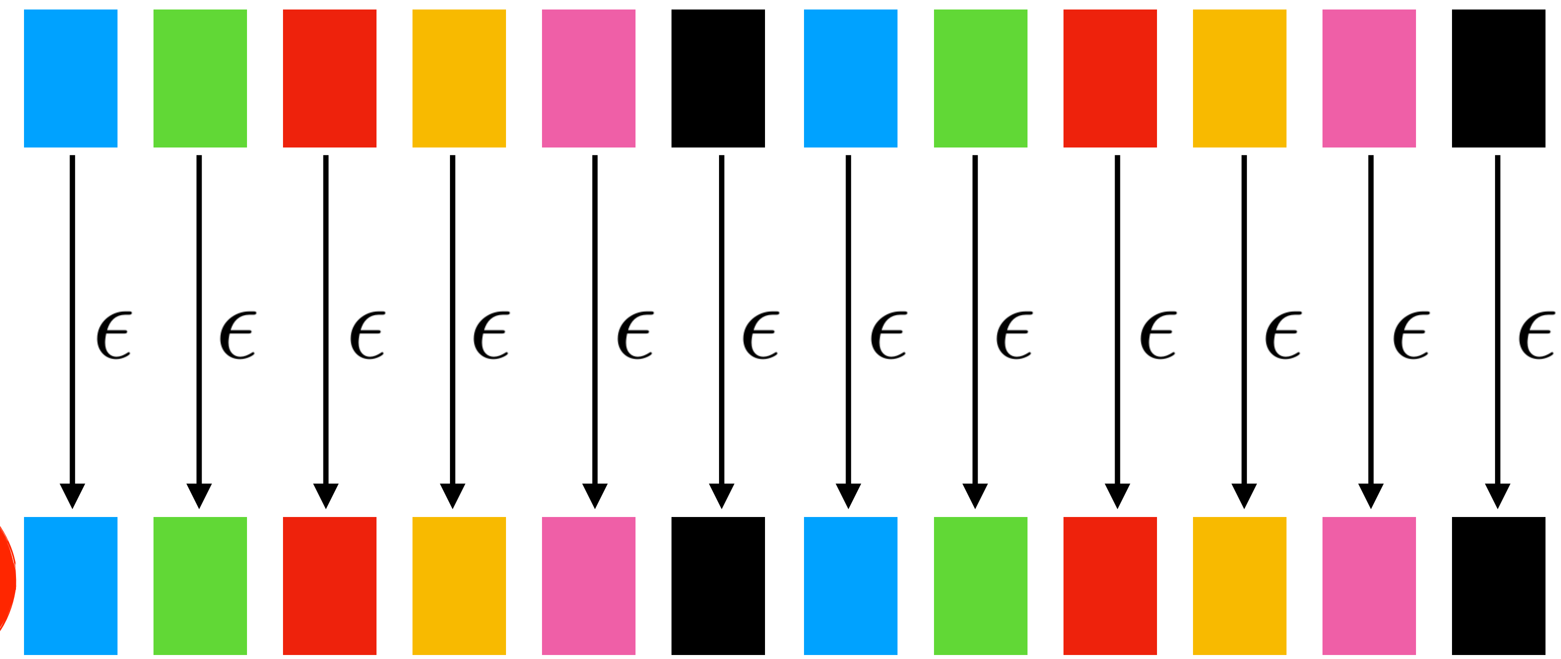


Romeo Kienzler

$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

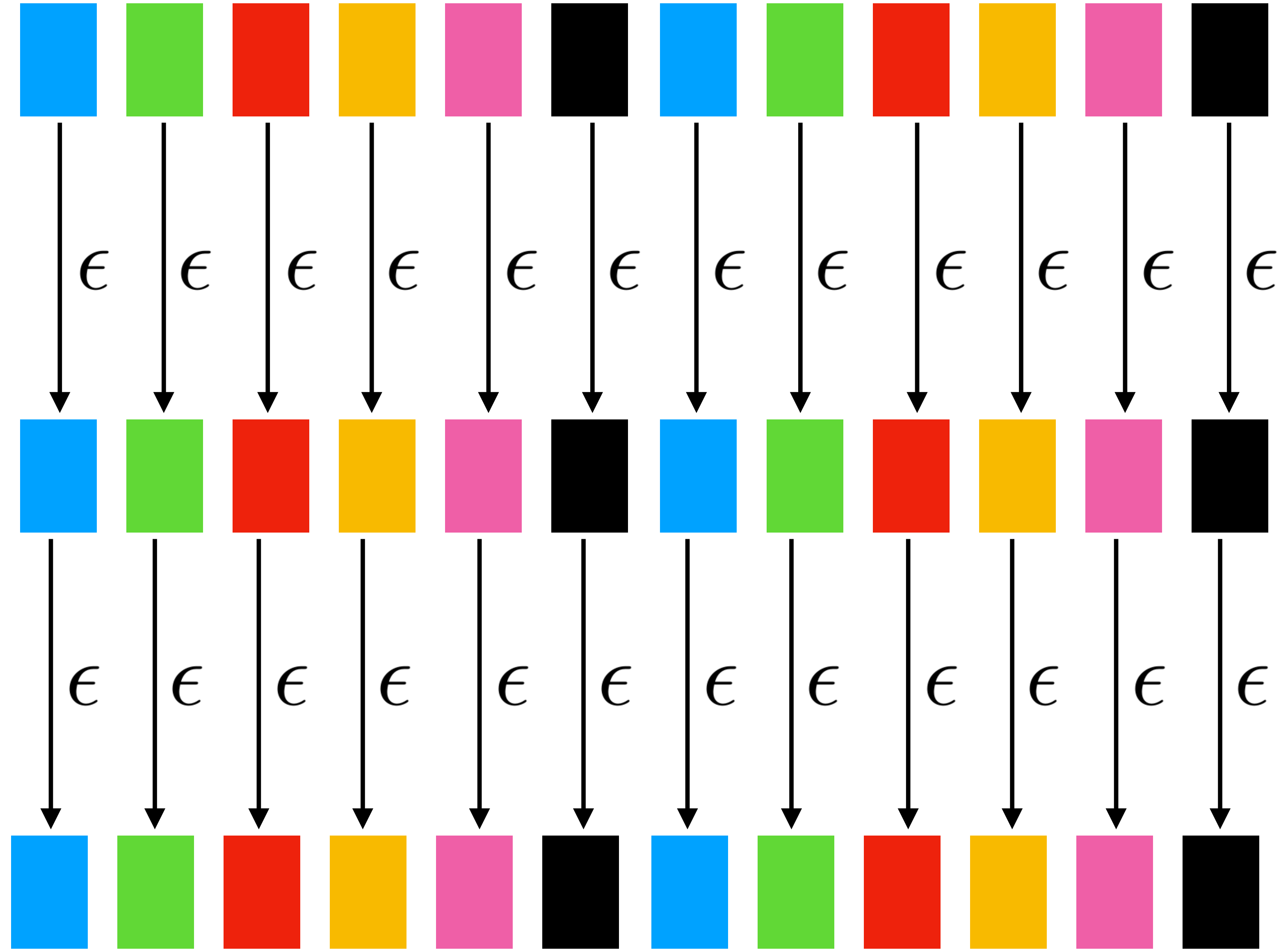


$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

$$\epsilon_2 = y - F_1(X)$$



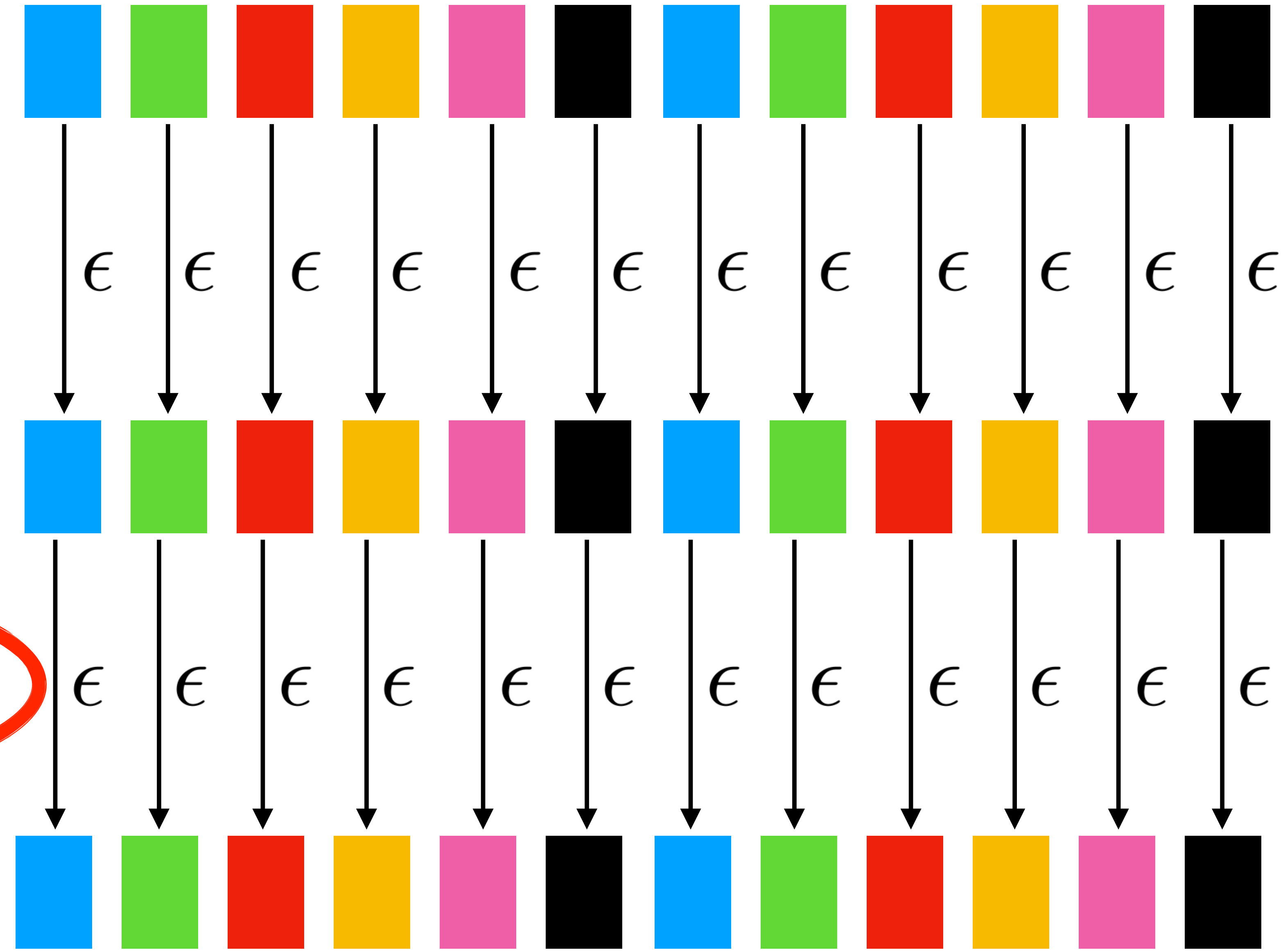
 Romeo Kienzler

$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

$$\epsilon_2 = y - F_1(X)$$



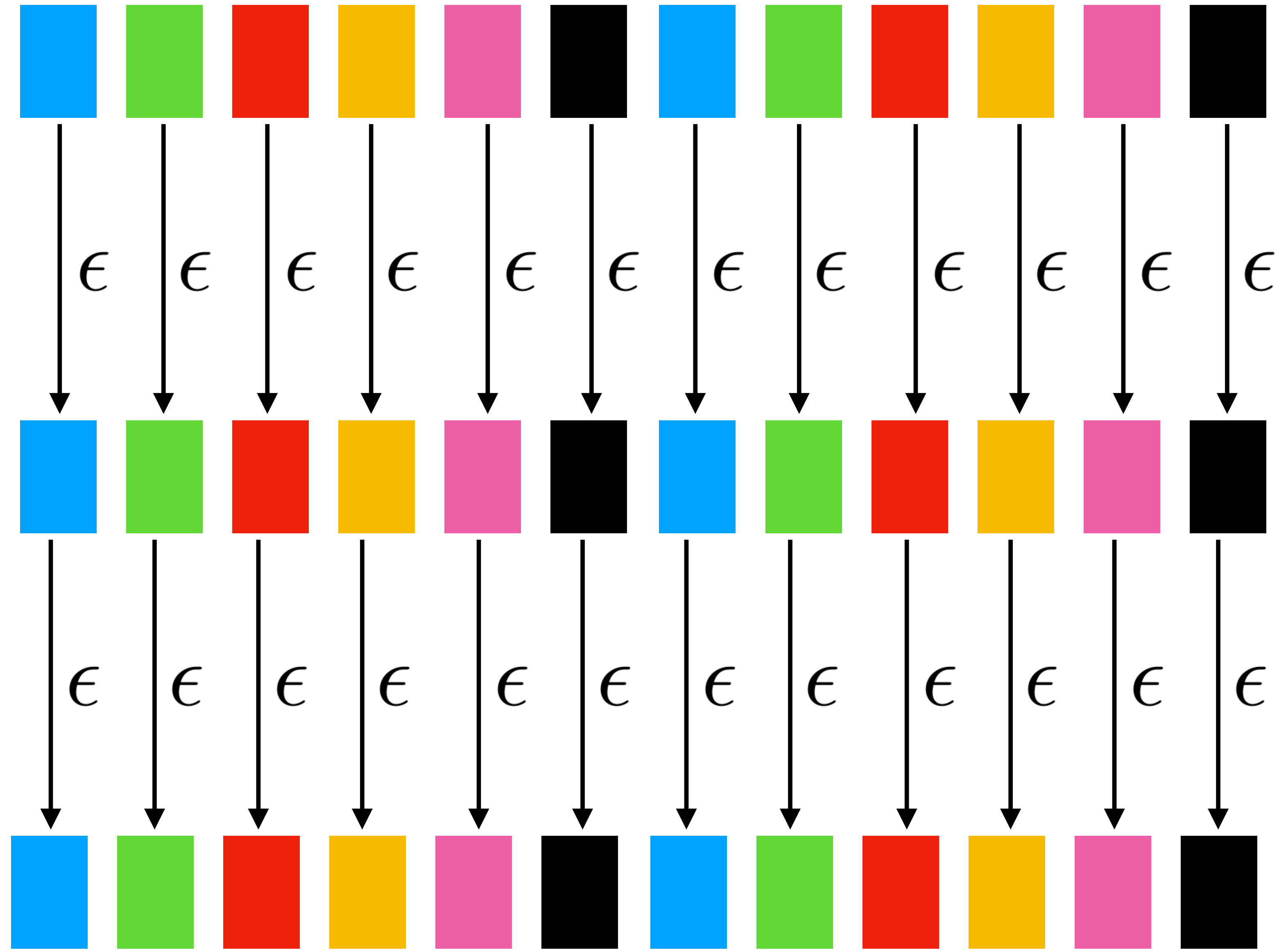
$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

$$\epsilon_2 = y - F_1(X)$$

$$F_2(X) = F_1(X) + h_2(\epsilon_2)$$



 Romeo Kienzler

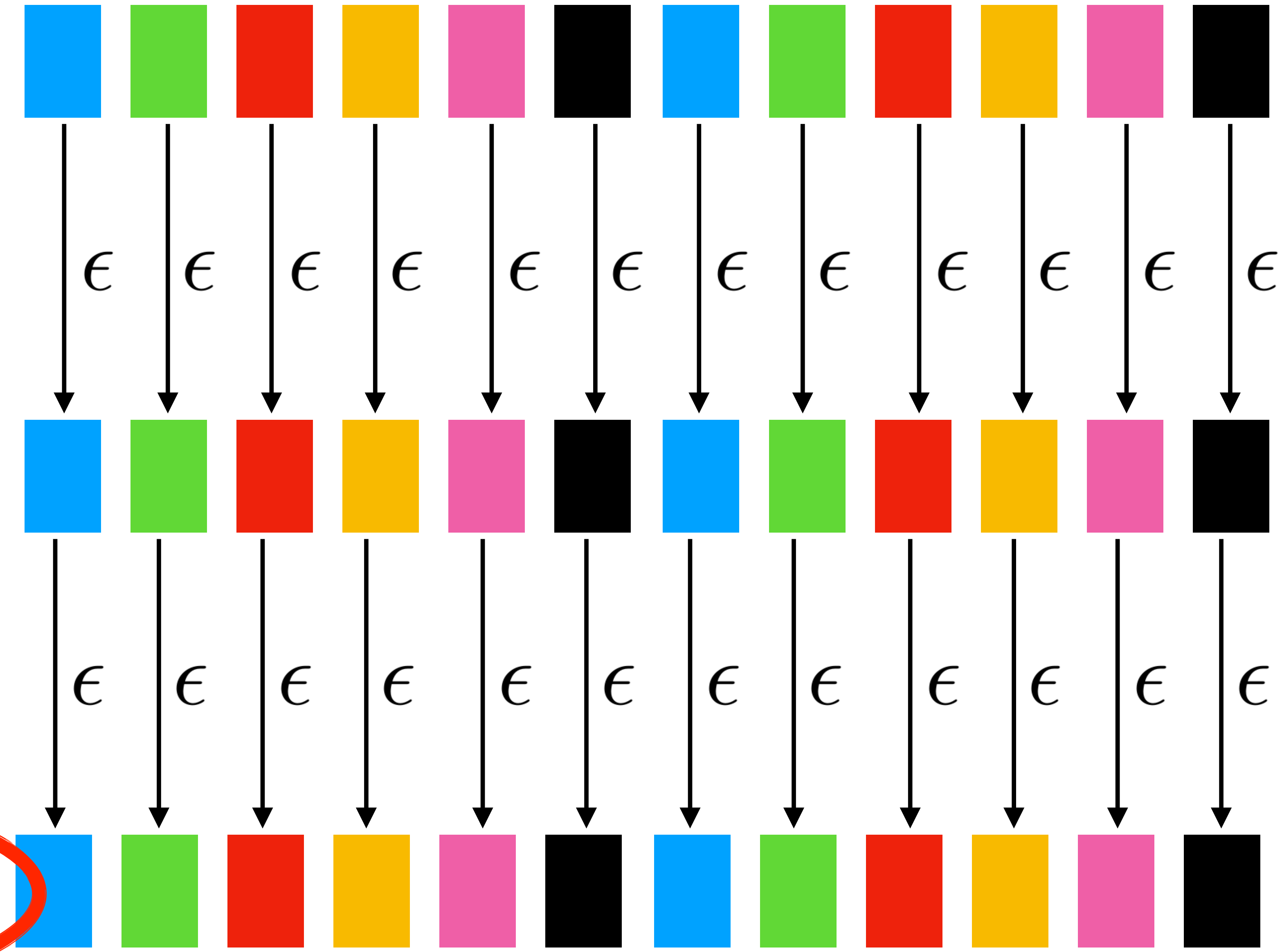
$$F_0(X) = h_0 = mean(y)$$

$$\epsilon_1 = y - F_0(X)$$

$$F_1(X) = F_0(X) + h_1(\epsilon_1)$$

$$\epsilon_2 = y - F_1(X)$$

$$F_2(X) = F_1(X) + h_2(\epsilon_2)$$



- + outperform RandomForest

- - computationally expensive (sequential nature)

XGBoost

Summary