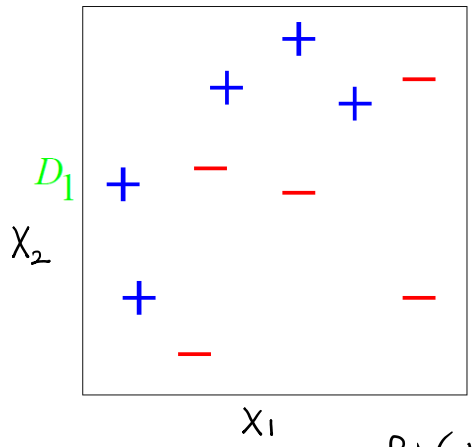
Boosting

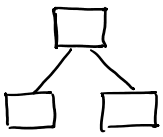
Saturday, July 15, 2023 8:15 AM

sns.scatterplot(data = ___, x = 'X1', y = 'X2', hue = 'y')

X_1	X_2	y
		+
		−

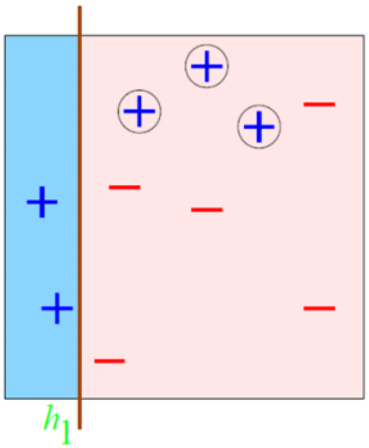


Stump

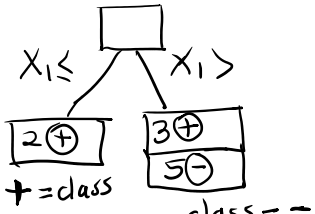


max_depth = 1

- fit(train)
- predict(train)

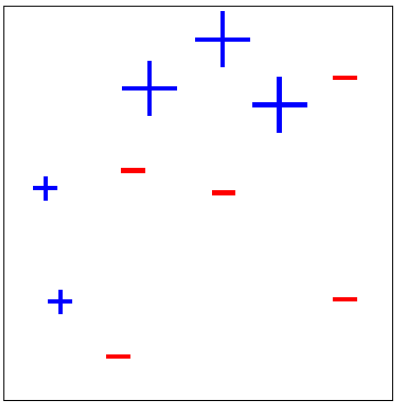


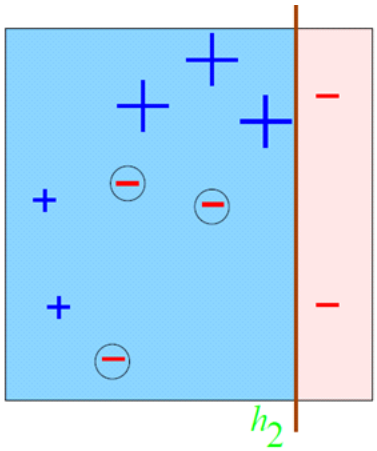
$X_1 < h_1$ $X_1 > h_1$



$\varepsilon_1 = 0.30$

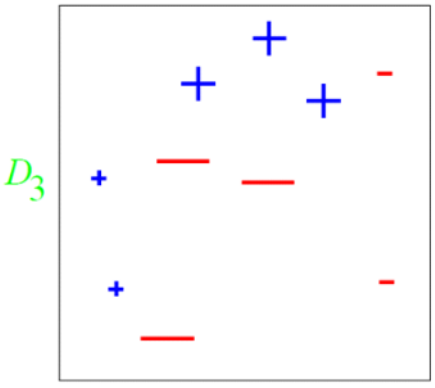
$\alpha_1 = 0.42$

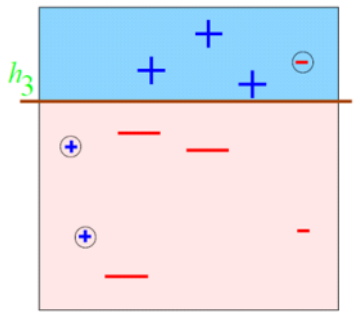




$\varepsilon_2 = 0.21$

$\alpha_2 = 0.65$





$\varepsilon_3 = 0.14$
$\alpha_3 = 0.92$

Gradient Boosting Method (GBM)

X_1 X_2 ... X_p	y	y_pred	$error_1$	y_err_1	$error_2$	y_err_2
	23	22.4	0.6	0.4	0.2	0.1
	44	41.2	3.8	3.2	0.6	0.5
	89	90.4	-1.4	-1.2	-0.2	-0.1
	52	51.2	0.8	0.5	0.3	0.1
	20	29	-9	-7	-2	-0.5

dtr_1.fit(X_train, y_train)
dtr_1.predict(X_train)

dtr_2.fit(X_train, error1)
dtr_2.predict(X_train)

dtr_3.fit(X_train, $error_2$)

.... dtr_{50}

eta

η : learning rate (0, 1)
$0 < \eta < 1$

Testing

X_1 X_2 ... X_p

dtr_1.predict() + η dtr_2.predict() + η dtr_3.predict()
+ ... + η dtr_{50}.predict()

= Prediction (Final)

Classification

X_1 X_2 .. X_p	y	Pred_proba	logloss
	1	0.4	
	0	0.3	
	0	0.3	
	1		
	0		

- Gradient Boosting (GBM)
  - Extreme Gradient Boosting (XGBoost)
  - Cat Boost – Yandex
  - Light GBM – Microsoft