

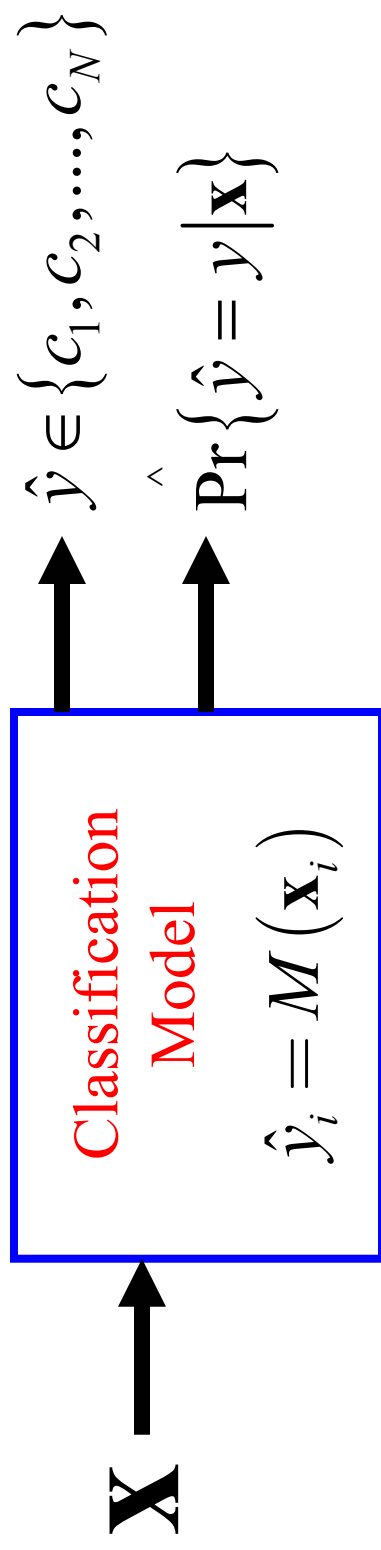
Class Project:
Artificial Intelligence 2:
Machine Learning
CSCI 4202

Greg Grudic

Probabilistic Classifiers

Given **training** data: $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$

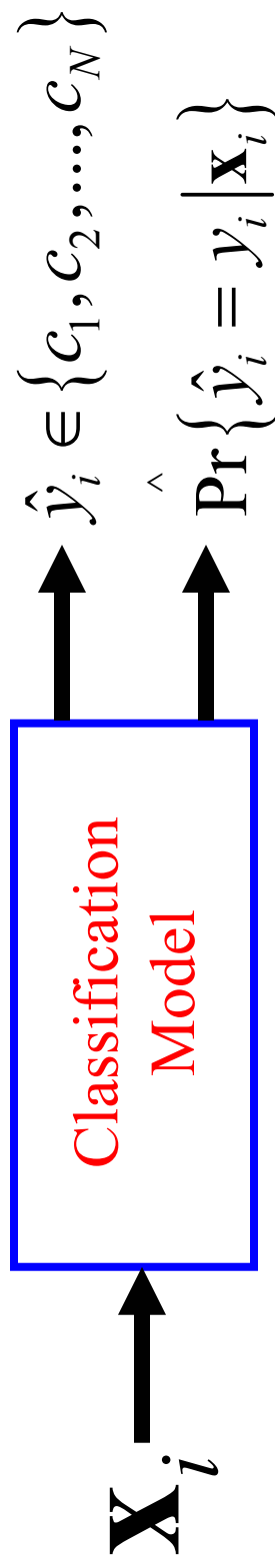
A model is computed:



How Good is a Probabilistic Classifier?

Given **TESTING** data: $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_K, y_K)$

Pass it through the model:



How good are the predictions \hat{y} and $\hat{\Pr}\{\hat{y} = y | \mathbf{x}\}$

What is the error rate on \hat{y} ?

$$\hat{y}_i = M(\mathbf{x}_i)$$

$$error_{M(\mathbf{x})} = \frac{1}{K} \sum_{i=1}^K c(\mathbf{x}_i, y_i, M(\mathbf{x}_i))$$

$$\text{Where } c(\mathbf{x}_i, y_i, M(\mathbf{x}_i)) = \begin{cases} 0 & \text{if } y_i = \hat{y}_i \\ 1 & \text{otherwise} \end{cases}$$

What is the error rate on

$$\Pr\{\hat{y} = y | \mathbf{x}\} ?$$

- We never see the actual probabilities on the accuracy of our \hat{y} class predictions!
- We never see the true $\Pr\{\hat{y} = y | \mathbf{x}\}$
- We only see $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_K, y_K)$
- **Your research goal is to find a way of estimating the error rate on $\Pr\{\hat{y} = y | \mathbf{x}\}$**

By Tuesday

- Write a program that takes as **input**

1. training data $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_K, y_K)$

2. Predictions $(\hat{y}_1, \Pr\{\hat{y}_1 = y_1 | \mathbf{x}_1\}), \dots, (\hat{y}_K, \Pr\{\hat{y}_K = y_K | \mathbf{x}_K\})$

1. And **outputs** an estimate of

$$\left\| \hat{\Pr}\{\hat{y}_1 = y_1 | \mathbf{x}_1\} - \Pr\{\hat{y}_1 = y_1 | \mathbf{x}_1\} \right\|$$