

# CSE 546 HW #2

Sam Kowash

October 19, 2018

## (1) A Taste of Learning Theory

1. Let  $X \in \mathbb{R}^d$  a random feature vector, and  $Y \in \{1, \dots, K\}$  a random label for  $K \in \mathbb{N}$  with joint distribution  $P_{XY}$ . We consider a randomized classifier  $\delta(x)$  which maps a value  $x \in \mathbb{R}^d$  to some  $y \in \{1, \dots, K\}$  with probability  $\alpha(x, y) \equiv P(\delta(x) = y)$  subject to  $\sum_{y=1}^K \alpha(x, y) = 1$  for all  $x$ . The risk of the classifier  $\delta$  is

$$R(\delta) \equiv \mathbb{E}_{XY} [\mathbf{1}\{\delta(X) \neq Y\}],$$

which we should interpret as the expected rate of misclassification. A classifier  $\delta$  is called deterministic if  $\alpha(x, y) \in \{0, 1\}$  for all  $x, y$ . Further, we call a classifier  $\delta_*$  a Bayes classifier if  $\delta_* \in \arg \inf_{\delta} R(\delta)$ .