

Population minimizers of different loss functions

loss function

$$\ell(y, f),$$

and risk

$$\mathbb{E}_{X,Y}[\ell(Y, f(X))].$$

population minimizer

$$f_{\text{pop}}(\cdot) = \operatorname{argmin}_{f(\cdot)} \mathbb{E}[\ell(Y, f(X))].$$

0-1 loss

$$f_{\text{pop}}(\cdot) = \text{Bayes classifier} \\ = \begin{cases} +1 & \text{if } \mathbb{P}[Y = +1|X = \cdot] > 1/2 \\ -1 & \text{if } \mathbb{P}[Y = +1|X = \cdot] < 1/2 \\ \text{undetermined} & \text{if } \mathbb{P}[Y = +1|X = \cdot] = 1/2 \end{cases}$$

SVM loss (hinge)

$$f_{\text{pop}}(\cdot) = \text{Bayes classifier}$$

squared error

$$f_{\text{pop}}(\cdot) = \mathbb{E}[Y|X = \cdot] = 2\mathbb{P}[Y = 1|X = \cdot] - 1$$

negative Bernoulli log-likelihood. $f_{\text{pop}}(\cdot) = \log\left(\frac{p(\cdot)}{1-p(\cdot)}\right)$

\rightsquigarrow Bernoulli log-likelihood and squared error loss yield probability estimates