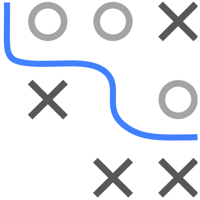
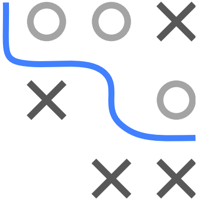
RISK MINIMIZER AND OPTIMAL CONSTANT



Name	Risk Minimizer	Optimal Constant
L2	$f^*(\mathbf{x}) = \mathbb{E}_{y\|x}[y \mid \mathbf{x}]$	$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
L1	$f^*(\mathbf{x}) = \text{med}_{y\|x}[y \mid \mathbf{x}]$	$\hat{f}(\mathbf{x}) = \text{med}(y^{(i)})$
0-1	$h^*(\mathbf{x}) = \arg\max_{l\in\mathcal{Y}} \mathbb{P}(y = l \mid \mathbf{x})$	$\hat{h}(\mathbf{x}) = \text{mode}\left\{y^{(i)}\right\}$
Brier	$\pi^*(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$	$\hat{\pi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
Bernoulli (on probs)	$\pi^*(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$	$\hat{\pi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
Bernoulli (on scores)	$f^*(\mathbf{x}) = \log\left(\frac{\mathbb{P}(y=1 \mid \mathbf{x})}{1-\mathbb{P}(y=1 \mid \mathbf{x})}\right)$	$\hat{f}(\mathbf{x}) = \log\frac{n_{+1}}{n_{-1}}$

We see: For regression, the RMs model the conditional expectation and median of the underlying distribution. This makes intuitive sense, depending on your concept of how to best estimate central location / how robust this location should be.

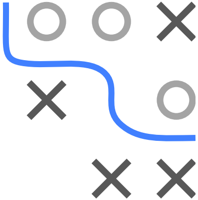
RISK MINIMIZER AND OPTIMAL CONSTANT



Name	Risk Minimizer	Optimal Constant
L2	$f^*(\mathbf{x}) = \mathbb{E}_{y\|x}\left[y \mid \mathbf{x}\right]$	$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
L1	$f^*(\mathbf{x}) = \text{med}_{y\|x}\left[y \mid \mathbf{x}\right]$	$\hat{f}(\mathbf{x}) = \text{med}(y^{(i)})$
0-1	$h^*(\mathbf{x}) = \arg\max_{l \in \mathcal{Y}} \mathbb{P}(y = l \mid \mathbf{x})$	$\hat{h}(\mathbf{x}) = \text{mode}\left\{y^{(i)}\right\}$
Brier	$\pi^*(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$	$\hat{\pi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
Bernoulli (on probs)	$\pi^*(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$	$\hat{\pi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
Bernoulli (on scores)	$f^*(\mathbf{x}) = \log\left(\frac{\mathbb{P}(y=1 \mid \mathbf{x})}{1-\mathbb{P}(y=1 \mid \mathbf{x})}\right)$	$\hat{f}(\mathbf{x}) = \log\frac{n_{+1}}{n_{-1}}$

For the 0-1 loss, the risk minimizer constructs the **optimal Bayes decision rule**: We predict the class with maximal posterior probability.

RISK MINIMIZER AND OPTIMAL CONSTANT



Name	Risk Minimizer	Optimal Constant
L2	$f^*(\mathbf{x}) = \mathbb{E}_{y\|x}[y \mid \mathbf{x}]$	$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
L1	$f^*(\mathbf{x}) = \text{med}_{y\|x}[y \mid \mathbf{x}]$	$\hat{f}(\mathbf{x}) = \text{med}(y^{(i)})$
0-1	$h^*(\mathbf{x}) = \arg\max_{l \in \mathcal{Y}} \mathbb{P}(y = l \mid \mathbf{x})$	$\hat{h}(\mathbf{x}) = \text{mode}\left\{y^{(i)}\right\}$
Brier	$\pi^*(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$	$\hat{\pi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
Bernoulli (on probs)	$\pi^*(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$	$\hat{\pi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$
Bernoulli (on scores)	$f^*(\mathbf{x}) = \log\left(\frac{\mathbb{P}(y=1 \mid \mathbf{x})}{1-\mathbb{P}(y=1 \mid \mathbf{x})}\right)$	$\hat{f}(\mathbf{x}) = \log\frac{n_{+1}}{n_{-1}}$

For Brier and Bernoulli, we predict the posterior probabilities (of the true DGP!). Losses that have this desirable property are called **proper scoring (rules)**.