Introduction: Learning with rejection

```
Classified and rejector trained simultaneously
```

```
Rejection area
```

```
Corollary 5: Necessary condition for rejection calibration
```

```
Supremum and infimum values coincide under the same constraint.
```

```
Goal Minimize $R_{0-1,c}(r,f) = \max_{p \in \mathcal{P}(Y)} \left( \mathcal{L}_{0-1,c}(r,f; x, y) \right)$
```

```
L_{0-1,c}(r,f;x,y) = \frac{1}{p(y|x)} \mathbb{1}_{r(x)>0} + c \mathbb{1}_{r(x)\leq 0}
```

Multiclass classification with rejection

Given: Labeled data: $(x_i, y_i)_{i=1}^n \sim \mathcal{P}(X \times Y)$

```
Chow (1970); Ramaswamy (2018)
```

```
Find:
Classifier: $c(x) = \arg\max_{y \in Y} q_{\theta}(x)$
```

```
Rejection:
$r(x) \in \mathbb{R}$
```

Goal: Minimize $R_{0-1,c}(r,f) = \max_{p \in \mathcal{P}(Y)} \left( \mathcal{L}_{0-1,c}(r,f; x, y) \right)$

Multiclass classification with rejection

Calibration

```
Optimal solution of classification with rejection: $f^*(x) = \arg\max_{y \in Y} q_{\theta}(y|x)$
```

```
Chow (1970)
```

```
r^*(x) = \max_{y \in Y} q_{\theta}(y|x) - (1-c)
```

```
(r,f) is calibrated if $R_{0-1,c}(r,f) = R_{0-1,c}(r^*,f^*)$.
```

```
f^* is classification-calibrated if $f(x)^* = f^*(x)$.
```

```
r^* is rejection-calibrated if $\text{sign}(r(x)) = \text{sign}(r^*(x))$.
```

A minimizer of a surrogate loss should give a calibrated $(r,f)$. 

```
Class and rejector are trained simultaneously
```

```
Classifier and rejector
```

```
Rejection area
```

Confidence-based approach

```
Rejector depends solely on classifier's confidence
```

```
Cross-entropy (CE) loss:
```

```
One-versus-all (OVA) loss:
```

```
Rejector:
$r_j(x) = \max_j \Psi_j(g_j(x)) - (1-c)
```

We provide excess risk bounds to guarantee OVA and CE losses.

Excess risk:

```
\Delta R_{0-1,c}(r,f) = R_{0-1,c}(r,f) - f^* \text{measurable}
```

Excess risk bound of OVA loss:

```
(2C)^{-a} \Delta R_{0-1,c}(r,f) \leq \Delta R_{OVA}(f)
```

Excess risk bound of CE loss:

```
\frac{1}{2} \Delta R_{0-1,c}(r,f)^2 \leq \Delta R_{CE}(f)
```

Experiments

Classifier-rejector: MPC-log (MPC with logistic loss), APC-log (APC with logistic loss)

Confidence-based: OVA+hin by Ramaswamy (2018), OVA-log (OVA with logistic loss), CE 0-1-c error:

Accuracy of non-rejected data: "-c" indicates all data were rejected.

References