Learning bounds for Risk-sensitive Learning
JaeHo Lee, Sejun Park, Jinwoo Shin Korea Advanced Institute of Science and Technology (KAIST)

**TL;DR.** We formulate risk-averse/seeking learning algorithms as an empirical OCE minimization, and give theoretical generalization guarantees.

### Motivation. Robust/Fair ML algorithms

Robust/Fair ML algorithms discriminate samples, based on their losses.

**Fact.** ML algorithms can be viewed as "minimizing the weighted sum" of losses: Given the training data \( z_1, z_2, \ldots, z_n \), we find a parameter \( \theta \in \Theta \) (e.g. neural network weights) that achieve

\[
\min_{\theta \in \Theta} \sum_{i=1}^{n} w_i \cdot l(z_i; \theta)
\]

for some weights \( w_1, w_2, \ldots, w_n \).

The weights are typically:

- **Classical ML.** Every samples are treated equally important.
- **Robust ML.** High-loss samples are viewed as "outliers," and are disregarded or considered less important.
- **Fair ML.** Reducing the loss of high-loss samples is prioritized, to mitigate the sense of unfairness among individuals.

### Background. OCE... what’s that?

**Prop. 1.** (Informal) As a special case, this modification incorporates algorithms that ignore high-loss samples.

**Prop. 2.** (Informal) Inverted OCE risks are more robust than the average loss, in terms of the influence function.

#### Result#1. Inverting OCE for Robust ML

We newly define inverted OCE to formally address robust ML algorithms

**Inverse.** We propose the "inverted OCE" to characterize the robust-ML-like algorithms which disregard high-loss samples.

\[
\text{OCE}(\theta) \triangleq \sup_{\lambda \in \mathbb{R}} \left\{ \lambda + E_P [\phi(l(Z; \theta) - \lambda)] \right\}
\]

**Remark.** In proving Thm. 2., we observe that OCE minimization is almost equivalent to the sample variance penalization procedure.

Batch-based sample variance penalization is often less noisier than the batch-based OCE minimization (using full-batch information).

### Result#2. Performance guarantees

Rademacher complexity bounds for empirical OCE minimizers.

**EOM.** Similar to ERM (empirical risk minimization), we consider EOM procedure:

\[
\hat{\theta}_{\text{eom}} \triangleq \arg \min_{\theta \in \Theta} \text{OCE}(\theta, P_n)
\]

where \( P_n \) is the empirical distribution of the training dataset.

**Thm. 1.** (Informal) We have the excess OCE risk bound: with high probability,

\[
\text{OCE}(\hat{\theta}_{\text{eom}}, P) - \inf_{\theta \in \Theta} \text{OCE}(\theta, P) \leq \mathcal{O} \left( \frac{\text{Lip}(\phi) \cdot \text{comp}(\Theta)}{\sqrt{n}} \right)
\]

where \( \text{comp}(\Theta) \) denotes the Rademacher complexity of \( \Theta \).

**Thm. 2.** (Informal) We have the excess mean loss bound: with high probability,

\[
\mathbb{E}(\hat{\theta}_{\text{eom}}, P) - \inf_{\theta \in \Theta} \mathbb{E}(\theta, P) \leq \mathcal{O} \left( \frac{\text{comp}(\Theta)}{\sqrt{n}} \right) + \varepsilon
\]

where \( \varepsilon \) is a small term proportional to the loss standard deviation of the optimal hypothesis.

### Result#3. Algorithmic implications

Sample variance penalization can be used for OCE minimization, provably.

**Remark.** In proving Thm. 2., we observe that OCE minimization is almost equivalent to the sample variance penalization procedure.

Batch-based sample variance penalization is often less noisier than the batch-based OCE minimization (using full-batch information).

**Idea.** Why don’t we use sample variance penalization as a baseline method for the empirical OCE minimization?

**Result.** The proposed baseline outperforms naïve batch CVaR minimization!