Lecture 16

Decorrelated Classifiers

Suppose we have a binary classifier \( \hat{X} \rightarrow \hat{f}_{cl}(\hat{X}) \) that we want \( f_{cl} \) to be statistically independent of.

Suppose we have a "protected" feature \( m \) that we want \( f_{cl} \) to be statistically independent of.

How to accomplish this?

Cannot in general just "exclude" \( m \) from training features \( \hat{X} \).

Motivation: want use \( m \) for sideband by estimation.

Before classifier

\( \hat{f}_{cl}(\hat{X}) \)

\( m \)

\( \hat{m}_{top} \)

\( \hat{m}_{bot} \)
Metrics for decorrelation: 

\[ JSO(p, q) = \frac{1}{2} \left( KL(p \| \frac{p+q}{2}) + KL(q \| \frac{p+q}{2}) \right) \]

\[ KL(p \| m) = \int dx \, p(x) \log \frac{p(x)}{m(x)}. \]

Measure of similarity between two distributions

\[ 0 \leq JSO \leq 1 \leftarrow \text{if } JSO \text{ computed by } \log_2 p \]

\[ p = q \quad \text{more similar} \quad p \text{ and } q \text{ no overlap at all} \]

Choose a cut \( \rightarrow \) e.g., cut at 50% signal efficiency. JSO50.

\( \text{Fpr} \cdot \text{Tpr} = R50 \)

\[ R50 \rightarrow \] more decorrelation

More classification