CSCI 6968 Weekly Participation 1

Shiuli Subhra Ghosh

January 20, 2023

1 Question

Often we can have a high number of potential features to be used in predicting our target y, e.g. $x \in R^{1000}$, and a large number of these features may not be relevant to the prediction of the target.

1. Let S be indices of some subset of the features, and x_S denote the corresponding random vector. Use the notion of independence to explain when the features x_S are irrelevant to predicting y.
2. More subtly, if we have a good subset of predictors x_G already, then we may say that a candidate set of features x_S doesn't add any additional value on top of x_G in predicting y. Use the notion of conditional independence to explain when this happens.

Answer(1): Let S be indices of some subset of the features, and x_S denote the corresponding random vector. When y is independent of x_S features, then it will be futile to predict y using x_S. Mathematically, when $y \perp\!\!\!\perp x_S$ or $P(y \mid x_S) = P(y)$.

Alternatively, we can say that if out of all the x features, the set x_S is not relevant to predict y, and x'_S is relevant to predict y, such that $x_S \cup x'_S = x$. Then we can say that y is independent of x_S given x'_S, $y \perp\!\!\!\perp x_S \mid x'_S$ or $P(y \mid \{x_S, x'_S\}) = P(y \mid x'_S)$.

Answer(2): More subtly, When we already have some good predictors x_G, then these x_S features will be irrelevant in predicting y given x_G. Mathematically writing, when $y \perp\!\!\!\perp x_S \mid x_G$, then features in x_S are irrelevant in predicting y.