WEEKLY PARTICIPATION 1: CONDITIONAL INDEPENDENCE

Xinshi Wang 661975305

January 2023

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

(a). 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.

When the features X_s are irrelevant to predicting y, we have the conditional independence $P(y|X_s) = P(y)$.

(b). More subtly, if we have a good subset of predictors $\mathbf{x_G}$ already, then we may say that a candidate set of features $\mathbf{x_S}$ doesn't add any additional value on top of $\mathbf{x_G}$ in predicting $\mathbf{y}$. Use the notion of conditional independence to explain when this happens.

When a candidate set of features X_S doesn't add any additional value on top of X_G in predicting y, we have $y \perp\!\!\!\perp X_S|X_G$, i.e., $P(y, X_S|X_G) = P(y|X_G)P(X_S|X_G)$.

The definition of a convex function can be written in LaTeX as:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

And the transformation for a nonnegative multiple of a convex function can be written as:

$$g(\lambda x_1+(1-\lambda)x_2) = c\cdot f(\lambda x_1+(1-\lambda)x_2) \leq c\cdot[\lambda f(x_1)+(1-\lambda)f(x_2)] = \lambda g(x_1)+(1-\lambda)g(x_2)$$