WEEKLY PARTICIPATION 1: CONDITIONAL INDEPENDENCE

Often we can have a high number of potential features to be used in predicting our target $\mathbf{y}$, e.g. $\mathbf{x} \in \mathbb{R}^{1000}$, 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 $\mathbf{x}_S$ denote the corresponding random vector. Use the notion of independence to explain when the features $\mathbf{x}_S$ are irrelevant to predicting $\mathbf{y}$.

(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.