WEEKLY PARTICIPATION 1: CONDITIONAL INDEPENDENCE

Recall that random vectors $\mathbf{x}$ and $\mathbf{y}$ are conditionally independent given $\mathbf{z}$, written $\mathbf{x} \perp\!\!\!\perp \mathbf{y} | \mathbf{z}$, if the conditional joint pmf/pdf factorizes into the product of the conditional marginals:

$$p(\mathbf{x}, \mathbf{y} | \mathbf{z}) = p(\mathbf{x} | \mathbf{z}) p(\mathbf{y} | \mathbf{z}).$$

Show that if $\mathbf{x} \perp\!\!\!\perp \mathbf{y} | \mathbf{z}$ then in fact we also have

$$p(\mathbf{y} | \mathbf{x}, \mathbf{z}) = p(\mathbf{y} | \mathbf{z}).$$

This follows from simple algebraic manipulations using the definitions of conditional pmfs/pdfs.

Comments.

(1) This result implies that if we are trying to predict $\mathbf{y}$ from both $\mathbf{x}$ and $\mathbf{z}$, then in fact we can do just as well by completely ignoring $\mathbf{x}$.

Using the example given in class, if we want to predict a student's first salary upon graduation, $\mathbf{y}$, and we know their gpa, $\mathbf{z}$, and their IQ, $\mathbf{x}$, then assuming that $\mathbf{y} \perp\!\!\!\perp \mathbf{x} | \mathbf{z}$, we can get just as accurate a prediction using only their GPA.

(2) We mentioned feature selection. One way to go about feature selection is, given a target r.v. $\mathbf{y}$ and a large set of features $\mathbf{x}$, to try to partition the features to identify a small set of relevant features, $\mathbf{x} = (\mathbf{x}_{\text{irrelevant}}, \mathbf{x}_{\text{relevant}})$, where relevance means $\mathbf{y} \perp\!\!\!\perp \mathbf{x}_{\text{irrelevant}} | \mathbf{x}_{\text{relevant}}$. This then allows you to fit more stable, simpler, and meaningful models, and can save a lot of computation.