On the Information Bottleneck

Blah

1 Information Bottleneck

Let random variable X denote an input source, Z a compressed representation, and Y observed output. We assume a Markov chain $Y \leftrightarrow X \leftrightarrow Z$. That is, Z cannot directly depend on Y. Then, the joint distribution $p(X, Y, Z)$ factorizes as

$$p(X, Y, Z) = p(Z|X, Y)p(Y|X)p(X) = p(Z|X)p(Y|X)p(X). \quad (1)$$

where we assume $p(Z|X, Y) = p(Z|X)$.

Our goal is to learn an encoding Z that is maximally informative about our target Y. As a measure we use the mutual information $I(Z, Y)$ between our encoding and output

$$I(Z, Y) = \int \int p(z, y) \log \frac{p(z, y)}{p(z)p(y)} dy\, dz. \quad (2)$$

If this was our only objective, the trivial identity encoding ($Z = X$) would always ensure a maximal informative representation. Instead, we would like to find the maximally informative representation subject to a constraint on it's complexity. Naturally, we would like to constraint the mutual information between our encoding Z and the input data Z such that $I(X, Z) \leq I_c$ where I_c denotes the information constraint. This suggests our objective:

$$\min_{P(Z|X)} I(X, Z) \quad \text{s.t.} \quad I(Z, Y) \leq I_c. \quad (3)$$

Our goal is to learn an encoding Z that is maximally expressive about Y while being maximally compressive about X.

References