

# Rewriting Mutual Information (MI) Using KL-divergence

The observed joint distribution of  $X_{w1}$  and  $X_{w2}$



$$I(X_{w1}; X_{w2}) = \sum_{u \in \{0,1\}} \sum_{v \in \{0,1\}} p(X_{w1} = u, X_{w2} = v) \log_2 \frac{p(X_{w1} = u, X_{w2} = v)}{p(X_{w1} = u)p(X_{w2} = v)}$$



The expected joint distribution of  $X_{w1}$  and  $X_{w2}$   
if  $X_{w1}$  and  $X_{w2}$  were independent

MI measures the divergence of the actual joint distribution from the expected distribution under the independence assumption. The larger the divergence is, the higher the MI would be.