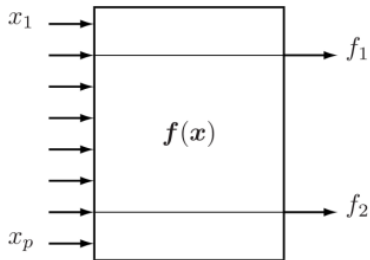


# Feature Selection

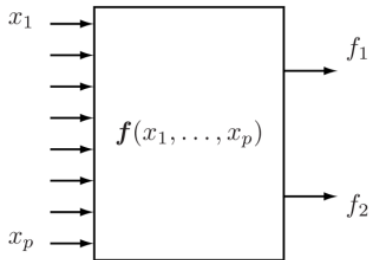
Clayton W. Seitz

February 21, 2022

# Feature Selection and Extraction



(a) feature selector



(b) feature extractor

# Feature Selection

What is it?

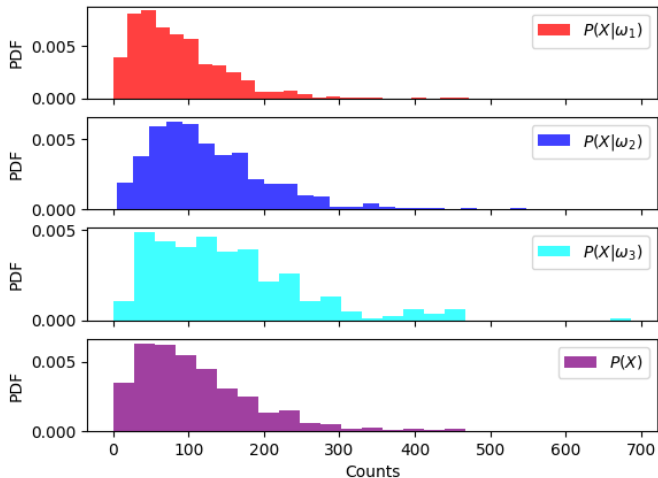
A special type of dimensionality reduction where we select a subset of features, in contrast with feature extraction like in PCA, VAE

Why do we do it?

- ▶ Quality of the input data is just as important as the algorithm you choose
- ▶ The volume of a feature space grows exponentially in the number of dimensions  $n$
- ▶ But we often have a small number of samples  $p \sim n$

# Probabilistic Distance Measures

How do we define a notion of distance for probability distributions?



# Symmetric Kullbeck-Leibler (KL) Divergence

The standard definition of KL-Divergence:

$$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$$

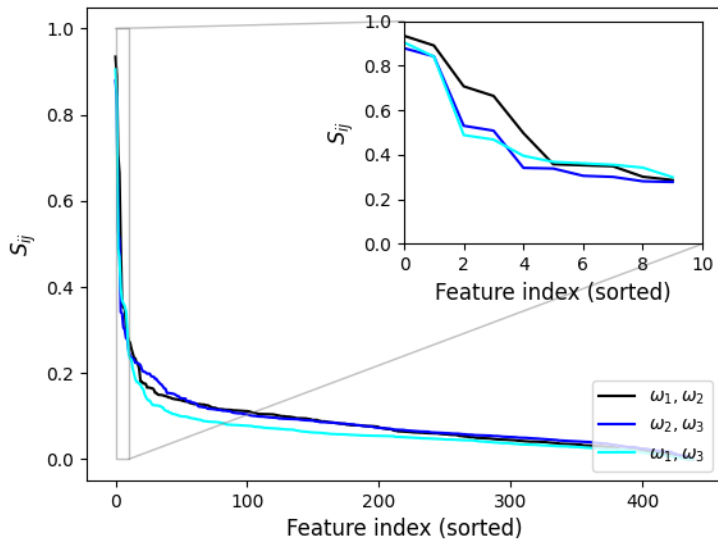
This is asymmetric. The form used in feature selection for classification tasks reads

$$S_{ij} = D_{KL}(P_i||Q_j) + D_{KL}(Q_j||P_i)$$

where  $P_i = P(X|\omega_i)$ ,  $Q_j = Q(X|\omega_j)$  for arbitrary classes  $\omega_i, \omega_j$

A major drawback of considering only conditional distributions is that we neglect pairwise or higher order changes

# Result on T1D Dataset



# References I