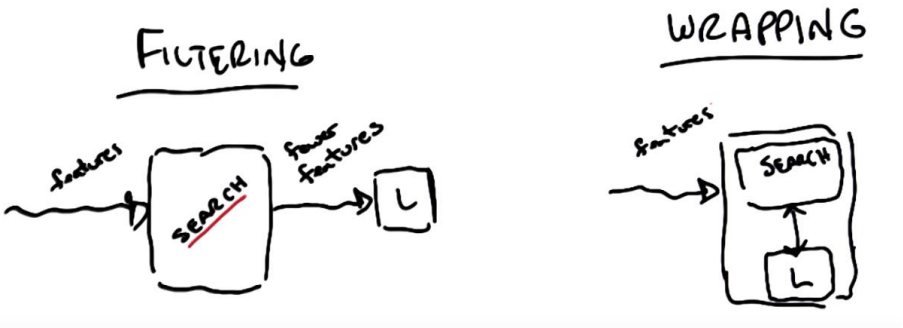
UL03. Feature Selection

Scan me

The Feature Selection Problem:

- Determine which features really matter.
- Understand the data better.
- Selecting the most relevant features helps in avoiding the Curse of Dimensionality.
- For N features, it takes exponential time (2^N) to select the most relevant subset of features M, where $M \leq N$. We have to try all the possible subsets of N to produce M.

Approaches to Feature Selection:



- Filtering: Given a set of features, apply a search algorithm to produce fewer features to be passed to the learning algorithm.
  - Faster than wrapping.
  - Doesn't account for the relationships between features.
  - Ignores the learning process (no feedback).
  - We can use an algorithm that can select the best features, like Decision Tree, then use another algorithm, with another inductive bias, for learning.
  - We can use different criterion to evaluate the usefulness of a subset of features:
    1. Information gain.
    2. Variance.
    3. Entropy.
    4. Eliminate dependent features.
- Wrapping: Given a set of features, apply a search algorithm to produce fewer features, pass them to the learning algorithm, then use the output to update the selected set of features.
  - Takes into account the learning model's bias.
  - Very slow.

- We can use different techniques to search:
  1. Randomized Optimization algorithms.
  2. Forward Selection: Select a feature and evaluate the effect of creating different combinations of this feature with other features. Stop once you stagnate. This is similar to Hill Climbing.
  3. Backward Elimination: Start with all the features and evaluate the effect of eliminating each feature.

Relevance:

- x_i is strongly relevant if removing it degrades the Bayes Optimal Classifier.
- x_i is weakly relevant if:
  - Not strongly relevant.
  - There's a subset of features S such that adding x_i to S improves the Bayes Optimal Classifier.
- x_i is irrelevant otherwise.
- Relevance vs. Usefulness:
  - Relevance measures the effect on B.O.C. This basically measures how much information a particular feature provides.
  - Usefulness measures effect on error. This basically measures if a feature helps in minimizing error given a particular model.