Active Model Selection

Johannes Kulick, M.Sc.
Universität Stuttgart
Machine Learning and Robotics Lab
Active Learning vs. Model Selection dilemma
Active Learning vs. Model Selection dilemma

Choose point that can be explained worst by the data so far
Active Learning vs. *Model Selection* dilemma

Given all other data points, how good can the model predict this one?
Active Model Selection

How to choose points for learning, such that they discriminate best between models?
Entropy minimization

First idea: minimize model posterior entropy!
(recall $H[X] = -\sum_X p(X) \log(p(X))$)
Entropy minimization

First idea: minimize model posterior entropy!

Model:

- a
- b
- c
- d (blue)
- e
Compare distributions instead

Model:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
</table>

How would the posterior change, if I would choose this point?
Compare distribution instead

\[ X^* = \arg \max_{X} \int p(Y|X, D) \cdot \Delta[p(M|D), p(M|D, X, Y)]dY \]
Compare distribution instead

\[ X^* = \arg\max_X \int p(Y|X, D) \cdot \Delta[p(M|D), p(M|D, X, Y)] dY \]

\[ \Delta[p, q] \text{ might be:} \]

- Change in Entropy:
  \[ \Delta_H[p, q] = |H[p] - H[q]| \]
Compare distribution instead

\[
X^* = \arg\max_X \int p(Y|X, D) \cdot \Delta[p(M|D), p(M|D, X, Y)]dY
\]

\(\Delta[p, q]\) might be:

- **Change in Entropy:**
  \[
  \Delta_H(p, q) = |H[p] - H[q]|
  \]

- **Kullback-Leibler-Divergence:**
  \[
  D_{KL}(p, q) = \sum_X p(X) \log \left( \frac{p(X)}{q(X)} \right)
  \]
Results on synthetic data
But in the end accuracy of prediction counts!