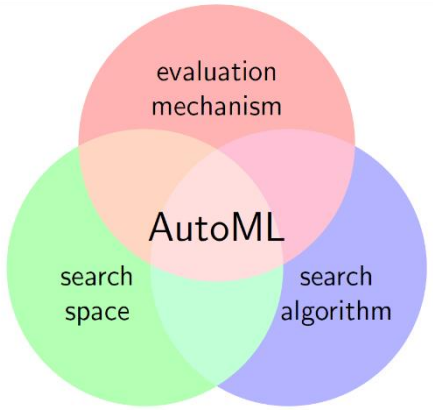
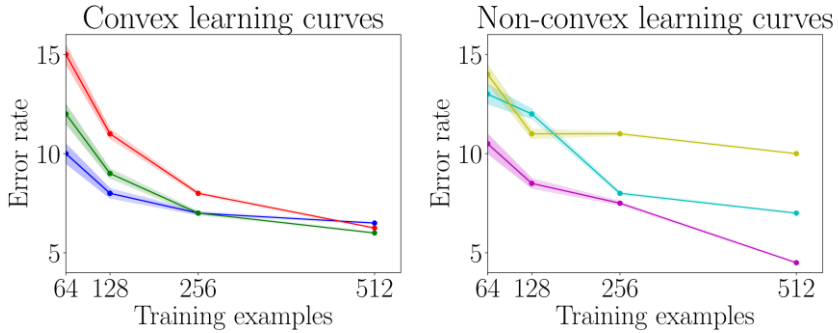
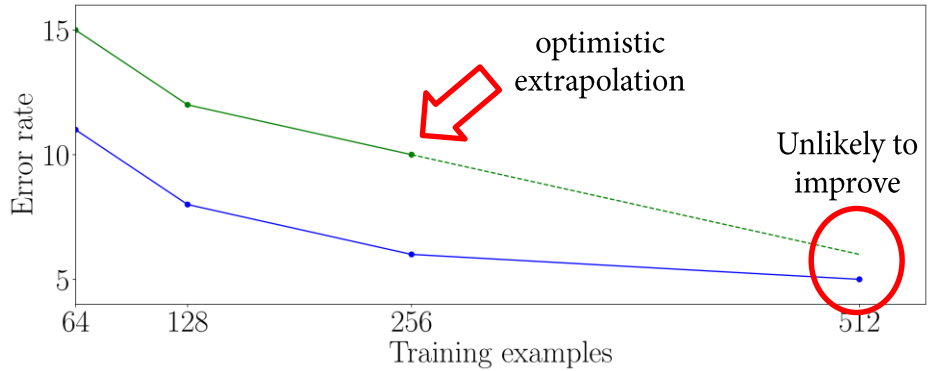
Towards Model Selection using Learning Curve Cross-Validation

Felix Mohr[1], Jan N. van Rijn[2]
[1]Universidad de La Sabana
[2]Leiden University





Assumption: Learning Curves are convex



Algorithm 1: LCCV: LearningCurveCrossValidation

$(s_1, .., s_T) \leftarrow$ initialize anchor points according to min_exp and data size;
$(C_1, .., C_T) \leftarrow$ initialize confidence intervals as $[0, 1]$ each;
$t \leftarrow 1$;
while $t \leq T \wedge (\sup C_T - \inf C_T > \varepsilon) \wedge |O_T| < n$ do
  repair_convexity $\leftarrow$ false;
  /* work in stage t: gather samples at current anchor point s_t */
  while $\sup C_t - \inf C_t > \varepsilon \wedge |O_t| < n \wedge \neg$ *repair_convexity* do
    add sample for s_t training points to O_t;
    update confidence interval C_t;
    if $t > 1$ then $\sigma_{t-1} = (\sup C_{t-1} - \inf C_t)/(s_{t-1} - s_t)$;
    if $t > 2 \wedge \sigma_{t-1} < \sigma_{t-2} \wedge |O_{t-1}| < n$ then
      repair_convexity $\leftarrow$ true;
  /* Decide how to proceed from this anchor point */
  if *repair_convexity* then
    $t \leftarrow t - 1$;
  else if *projected bound for* s_T *is* $> r + \delta$ then
    return $\perp$
  else if $r = 1 \vee (t \geq 3 \wedge$ IPL_ESTIMATE$(s_T) \leq r)$ then
    $t \leftarrow T$;
  else
    $t \leftarrow t + 1$
return $\langle mean(C_T), (C_1, .., C_T) \rangle$

Future Work: Can we predict whether two learning curves cross?