THE UNIVERSITY of EDINBURGH
informatics

Applied Machine Learning (AML)

Class Starting at 4:10pm

Oisin Mac Aodha • Siddharth N.

Applied Machine Learning

Week 6: Optimisation and Generalisation

This slides will be made available on the project website after the class.
This session will be recorded.

Overview

1) Outline your tasks this for week
2) Discussion of Week 5’s topics

Week 6: Your tasks for this week

1) Complete Tutorial 2 - solutions will be released later this week
2) Watch videos for week 6 - **Evaluation** and **Model Selection**
3) **Ask questions** on Piazza if stuck
4) Continue working on the coursework
5) Start Lab 3 which takes places next week - link in week 7

Coursework - Progress Report

- **Not** mandatory or assessed; but **highly** recommended
- A ‘check-in’ to see how everyone is doing
- Report - Wed 30 Oct 5pm
  - min 1-page interim report
  - include a section each on ‘Current Progress’ and ‘Plans for Completion’
- Feedback - Fri 01 Nov 1-3pm [AT 5.04]
  - intention **not** for detailed individual feedback; coarse level
  - primarily to identify those not yet started or doing very wrong things!
  - collective feedback to class after

Coursework Discussion on Piazza

- Please mark questions about coursework project as private.
- Ensure that the question is visible by all the instructors.
- Potential for too much discussion for each group to keep track of.
- We will compile all the relevant feedback into an FAQ on the course website—updated as and when relevant.

Wooclap

Bias and Variance

Expected Target Error

Targets sampled as $y \sim p_{\mathcal{D}}(y|\boldsymbol{x})$.

$$
\begin{aligned}
\mathbb{E}\left[(\hat{y}-y)^2|\boldsymbol{x}\right] &= \mathbb{E}\left[\hat{y}^2 - 2\hat{y}y + y^2|\boldsymbol{x}\right] \\
&= \hat{y}^2 - 2\hat{y}\,\mathbb{E}[y|\boldsymbol{x}] + \mathbb{E}\left[y^2|\boldsymbol{x}\right] && \text{(linearity of expectation)} \\
&= \hat{y}^2 - 2\hat{y}\,\mathbb{E}[y|\boldsymbol{x}] + \mathbb{E}[y|\boldsymbol{x}]^2 + \mathrm{Var}[y|\boldsymbol{x}] && \text{(expression for variance)} \\
&= (\hat{y} - \mathbb{E}[y|\boldsymbol{x}])^2 + \mathrm{Var}[y|\boldsymbol{x}] \\
&\triangleq \underbrace{(\hat{y} - y_\star)^2}_{\text{residual}} + \underbrace{\mathrm{Var}[y|\boldsymbol{x}]}_{\text{Bayes error}}
\end{aligned}
$$

Bias and Variance

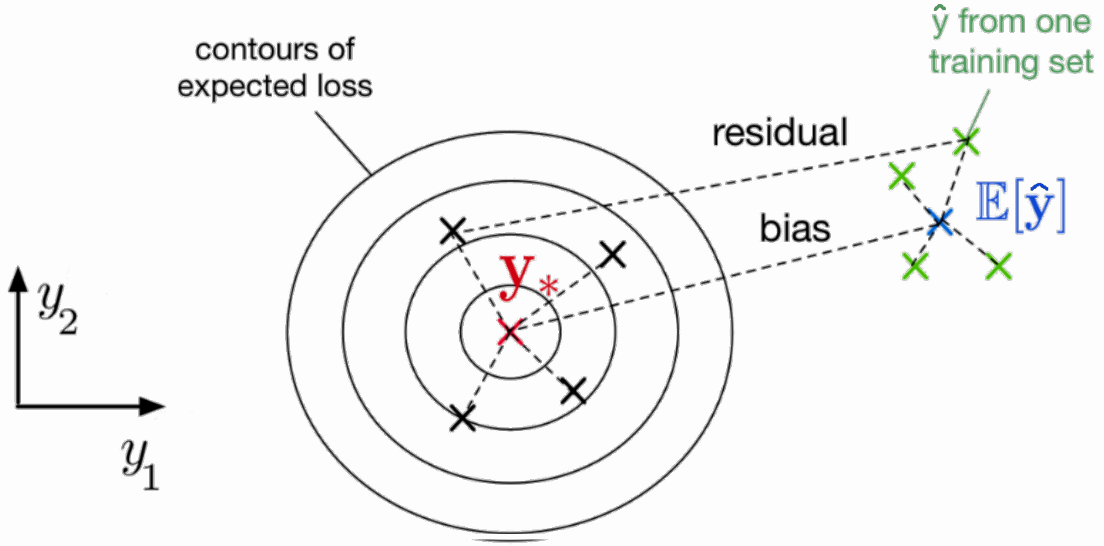
Expected Test Error

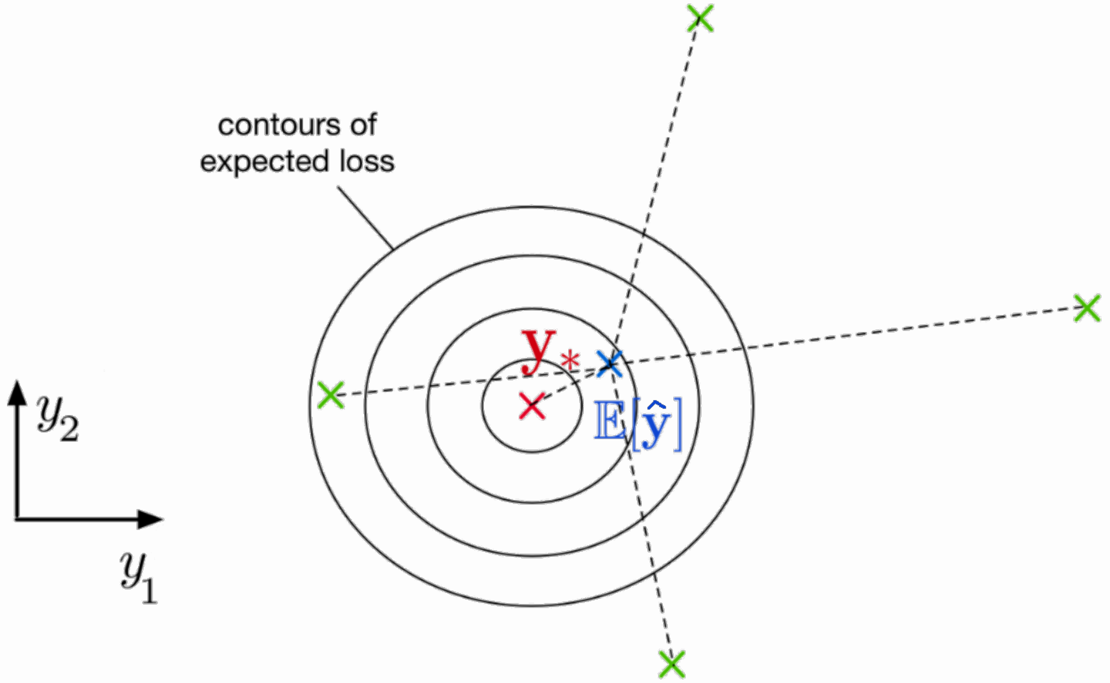
Assume model (p_w) trained on $\mathcal{D} \sim p_{\mathcal{D}}(\boldsymbol{x}, y)$; compute predictions on $\boldsymbol{x}$.

Predictions generated as $\hat{y} \sim p_{\boldsymbol{w}}(\hat{y}|\boldsymbol{x})$.

$$
\begin{aligned}
\mathbb{E}\left[(\hat{y} - y)^2\right] &= \mathbb{E}\left[(\hat{y} - y_\star)^2\right] + \mathrm{Var}[y] \\
&= \mathbb{E}\left[y_\star^2 - 2\hat{y}y_\star + \hat{y}^2\right] + \mathrm{Var}[y] \\
&= y_\star^2 - 2y_\star\,\mathbb{E}[\hat{y}] + \mathbb{E}\left[\hat{y}^2\right] + \mathrm{Var}[y] && \text{(linearity of expectation)} \\
&= y_\star^2 - 2y_\star\,\mathbb{E}[\hat{y}] + \mathbb{E}[\hat{y}]^2 + \mathrm{Var}[\hat{y}] + \mathrm{Var}[y] && \text{(expression for variance)} \\
&= \underbrace{(y_\star - \mathbb{E}[\hat{y}])^2}_{\text{bias}} + \underbrace{\mathrm{Var}[\hat{y}]}_{\text{variance}} + \underbrace{\mathrm{Var}[y]}_{\text{Bayes error}}
\end{aligned}
$$

Bias and Variance: Schematic





Generalisation Error:

average squared length of *residual* $\|\hat{y} - y\|^2$

Bias:

average squared length of *bias* $\|y_\star - \mathbb{E}[\hat{y}]\|^2$

Variance: spread of green $\times$'s

Bayes error: spread of black $\times$'s

Figures: Roger Grosse - Generalization