Foundations of Machine Learning

Introduction

Mahesan Niranjan

School of Electronics and Computer Science
University of Southampton

July 2021

Machine Learning: Data-driven Modelling

Data $\quad \{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N} \quad \{\boldsymbol{x}_n\}_{n=1}^{N}$

Machine Learning: Data-driven Modelling

Data $\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$ $\{\boldsymbol{x}_n\}_{n=1}^{N}$

Function Approximator $\boldsymbol{t} = f(\boldsymbol{x}, \boldsymbol{\theta}) + v$

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$	$\{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} \;=\; f\,(\boldsymbol{x},\,\boldsymbol{\theta}) \;+\; v$	
Parameter Estimation	$E_0 \;=\; \sum_{n=1}^{N} \{\lvert\lvert\, \boldsymbol{t}_n \;-\; f\,(\boldsymbol{x}_n;\,\boldsymbol{\theta})\;\rvert\rvert\}^2$	

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$	$\{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} \ = \ f\,(\boldsymbol{x},\, \boldsymbol{\theta}) \ + \ v$	
Parameter Estimation	$E_0 \ = \ \sum_{n=1}^{N} \,\{\|\|\, \boldsymbol{t}_n \ - \ f\,(\boldsymbol{x}_n;\, \boldsymbol{\theta})\ \|\|\}^2$	
Prediction	$\hat{\boldsymbol{t}}_{N+1} \ = \ f\left(\boldsymbol{x}_{N+1},\, \hat{\boldsymbol{\theta}}\right)$	

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$	$\{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} = f\,(\boldsymbol{x},\, \boldsymbol{\theta}) + v$	
Parameter Estimation	$E_0 = \sum_{n=1}^{N} \{\lvert\lvert\, \boldsymbol{t}_n - f\,(\boldsymbol{x}_n;\, \boldsymbol{\theta})\, \rvert\rvert\}^2$	
Prediction	$\hat{\boldsymbol{t}}_{N+1} = f\left(\boldsymbol{x}_{N+1},\, \hat{\boldsymbol{\theta}}\right)$	
Regularization	$E_1 = \sum_{n=1}^{N} \{\lvert\lvert\, \boldsymbol{t}_n - f\,(\boldsymbol{x}_n)\, \rvert\rvert\}^2 + r\,(\lvert\lvert\boldsymbol{\theta}\rvert\rvert)$	

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$	$\{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} \ = \ f\,(\boldsymbol{x},\, \boldsymbol{\theta}) \, + \, v$	
Parameter Estimation	$E_0 \ = \ \sum_{n=1}^{N} \, \{ \lvert\lvert \, \boldsymbol{t}_n \ - \ f\,(\boldsymbol{x}_n;\, \boldsymbol{\theta}) \ \rvert\rvert \}^2$	
Prediction	$\hat{\boldsymbol{t}}_{N+1} \ = \ f\left(\boldsymbol{x}_{N+1},\, \hat{\boldsymbol{\theta}}\right)$	
Regularization	$E_1 \ = \ \sum_{n=1}^{N} \, \{ \lvert\lvert \, \boldsymbol{t}_n \ - \ f\,(\boldsymbol{x}_n) \ \rvert\rvert \}^2 \, + \, r\,(\lvert\lvert \boldsymbol{\theta} \rvert\rvert)$	
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \middle| \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}\right)$	

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$	$\{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} = f(\boldsymbol{x}, \boldsymbol{\theta}) + v$	
Parameter Estimation	$E_0 = \sum_{n=1}^{N} \{\|\| \boldsymbol{t}_n - f(\boldsymbol{x}_n; \boldsymbol{\theta}) \|\|\}^2$	
Prediction	$\hat{\boldsymbol{t}}_{N+1} = f\left(\boldsymbol{x}_{N+1}, \hat{\boldsymbol{\theta}}\right)$	
Regularization	$E_1 = \sum_{n=1}^{N} \{\|\| \boldsymbol{t}_n - f(\boldsymbol{x}_n) \|\|\}^2 + r(\|\|\boldsymbol{\theta}\|\|)$	
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \middle\| \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}\right)$	
Probabilistic Inference	$\boldsymbol{E}[g(\boldsymbol{\theta})] = \int g(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$	

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^N \qquad \{\boldsymbol{x}_n\}_{n=1}^N$
Function Approximator	$\boldsymbol{t} = f(\boldsymbol{x}, \boldsymbol{\theta}) + v$
Parameter Estimation	$E_0 = \sum_{n=1}^N \{\|\| \boldsymbol{t}_n - f(\boldsymbol{x}_n; \boldsymbol{\theta}) \|\|\}^2$
Prediction	$\hat{\boldsymbol{t}}_{N+1} = f\left(\boldsymbol{x}_{N+1}, \hat{\boldsymbol{\theta}}\right)$
Regularization	$E_1 = \sum_{n=1}^N \{\|\| \boldsymbol{t}_n - f(\boldsymbol{x}_n) \|\|\}^2 + r(\|\|\boldsymbol{\theta}\|\|)$
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \| \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^N\right)$
Probabilistic Inference	$\boldsymbol{E}[g(\boldsymbol{\theta})] = \int g(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$
Sequential Estimation	$\boldsymbol{\theta}(n-1\|n-1) \longrightarrow \boldsymbol{\theta}(n\|n-1) \longrightarrow \boldsymbol{\theta}(n\|n)$

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$ $\qquad \{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} = f(\boldsymbol{x}, \boldsymbol{\theta}) + v$
Parameter Estimation	$E_0 = \sum_{n=1}^{N} \{\vert\vert\, \boldsymbol{t}_n - f(\boldsymbol{x}_n; \boldsymbol{\theta})\, \vert\vert\}^2$
Prediction	$\hat{\boldsymbol{t}}_{N+1} = f\left(\boldsymbol{x}_{N+1}, \hat{\boldsymbol{\theta}}\right)$
Regularization	$E_1 = \sum_{n=1}^{N} \{\vert\vert\, \boldsymbol{t}_n - f(\boldsymbol{x}_n)\, \vert\vert\}^2 + r(\vert\vert\boldsymbol{\theta}\vert\vert)$
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \vert \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}\right)$
Probabilistic Inference	$\boldsymbol{E}[g(\boldsymbol{\theta})] = \int g(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$
Sequential Estimation	$\boldsymbol{\theta}(n-1 \vert n-1) \longrightarrow \boldsymbol{\theta}(n \vert n-1) \longrightarrow \boldsymbol{\theta}(n \vert n)$

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n, \; \boldsymbol{t}_n\}_{n=1}^{N} \qquad \{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} \; = \; f\,(\boldsymbol{x}, \, \boldsymbol{\theta}) \, + \, v$
Parameter Estimation	$E_0 \; = \; \sum_{n=1}^{N} \, \{\|\| \, \boldsymbol{t}_n \; - \; f\,(\boldsymbol{x}_n; \, \boldsymbol{\theta}) \; \|\|\}^2$
Prediction	$\hat{\boldsymbol{t}}_{N+1} \; = \; f\left(\boldsymbol{x}_{N+1}, \, \hat{\boldsymbol{\theta}}\right)$
Regularization	$E_1 \; = \; \sum_{n=1}^{N} \, \{\|\| \, \boldsymbol{t}_n \; - \; f\,(\boldsymbol{x}_n) \; \|\|\}^2 \, + \, r\,(\|\|\boldsymbol{\theta}\|\|)$
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \middle\| \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}\right)$
Probabilistic Inference	$\boldsymbol{E}\,[g\,(\boldsymbol{\theta})] = \int g\,(\boldsymbol{\theta})\,p\,(\boldsymbol{\theta})\,d\boldsymbol{\theta} \; = \frac{1}{N_s}\sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$
Sequential Estimation	$\boldsymbol{\theta}\,(n-1\|n-1) \longrightarrow \boldsymbol{\theta}\,(n\|n-1) \longrightarrow \boldsymbol{\theta}\,(n\|n)$ Kalman & Particle Filters;

Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N}$	$\{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} = f(\boldsymbol{x}, \boldsymbol{\theta}) + v$	
Parameter Estimation	$E_0 = \sum_{n=1}^{N} \{\|\| \boldsymbol{t}_n - f(\boldsymbol{x}_n; \boldsymbol{\theta}) \|\|\}^2$	
Prediction	$\hat{\boldsymbol{t}}_{N+1} = f\left(\boldsymbol{x}_{N+1}, \hat{\boldsymbol{\theta}}\right)$	
Regularization	$E_1 = \sum_{n=1}^{N} \{\|\| \boldsymbol{t}_n - f(\boldsymbol{x}_n) \|\|\}^2 + r(\|\|\boldsymbol{\theta}\|\|)$	
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \middle\| \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}\right)$	
Probabilistic Inference	$\boldsymbol{E}[g(\boldsymbol{\theta})] = \int g(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$	
Sequential Estimation	$\boldsymbol{\theta}(n-1 \mid n-1) \longrightarrow \boldsymbol{\theta}(n \mid n-1) \longrightarrow \boldsymbol{\theta}(n \mid n)$ Kalman & Particle Filters; Reinforcement Learning	