Advanced Machine Learning Subsidary Notes

Lecture 6: Boosting

Adam Prügel-Bennett

January 19, 2024

1 Keywords

- Boosting, AdaBoost, Gradient Boosting

2 Main Points

2.0.1 Boosting

- Boosting constructs a *strong learner* as a weighted sum of *weak learners*
- Adaboost
  - Used for binary decisions
  - Start with a set of weak learners, $\mathcal{W}$
  - Each weak learner $h_i(\boldsymbol{x})$ outputs ± 1
  - Greedily build the strong learner by adding $\alpha_t\, h_t(\boldsymbol{x})$ at iteration t
  - Uses an exponential "error" to choose the weak learner and α_t
  - Algorithm does the following
    - Define a weight, w_t^μ, for each training example $(\boldsymbol{x}^\mu, y^\mu)$
      - initially these are set to 1
      - Large weight implies the training example is poorly predicted
    - Choose the weak learner, h_t that fails only where prediction is good
      - it decides this by summing the weights of training examples where the weak learner makes an error
      - it choose the weak learner with the smallest sum
    - Choose the parameter α_t to minimises the error
  - Need to understand derivation and resulting algorithm (this is complicated)
- Gradient Boosting
  - Used on regression problems
  - Iterative algorithm where we learn a new weak learner that minimises the residual errors
  - Uses very small decision trees for regression
- Performance of Boosting
  - Can over-fit (use early stopping)
  - Only works for very simple weak-learners (strong learners will over-fit)
  - Can give state-of-the-art performance