

Handout — Ensembling: Random Forests & Gradient Boosting

Kurz (Seite 1) — Kernaussagen

- Ziel von Ensembling: mehrere schwache Modelle kombinieren, um stabilere, genauere Vorhersagen zu erhalten.
- Random Forest: Bagging + Entscheidungsbäume → Varianzreduktion.
- Gradient Boosting: sequentielles Anpassen an Residuen → Biasreduktion bei kontrollierter Varianz.

Wichtige Formeln & Konzepte (Seite 2)

- Vorhersage (Bagging): $f(x) = \frac{1}{M} \sum_{m=1..M} f_m(x)$
- Gradient Boosting (sketch): $F_{t+1}(x) = F_t(x) + * h_t(x)$, wobei h_t Regression auf Residuen ist.
- Feature importance: mittlere Abnahme der impurity / Permutation importance.

Beispiel & Visualisierungen (Seite 3)

- Datensatz: Iris / Breast Cancer — kurze Tabelle mit Metriken (Accuracy, Precision, Recall)
- Plot: Feature importances für RF und GB, Lernkurven (Training vs Validation)

Praktische Hinweise (Seite 4)

- Hyperparameter: `n_estimators`, `max_depth`, `learning_rate` (für GB).
- Überanpassung: frühes Stoppen / Regularisierung (`max_depth`, `subsample`).
- Einsatzempfehlung: RF für schnelle, robuste Baselines; GB für sorgfältig getunte, performante Modelle.

Quellen & weiterführende Links

- scikit-learn documentation: ensemble module
- Hastie, Tibshirani, Friedman — The Elements of Statistical Learning