<working document>

Estimating Black Hole Mass Using Machine Learning

623.504, 25S Artificial Intelligence & Machine Learning

Walter Telsnig
wtelsnig@edu.aau.at

Estimating Black Hole Mass Using Machine Learning

1. Problem Statement

Supermassive black holes are found at the centers of most galaxies, and their masses play a crucial role in understanding galaxy formation and evolution. Traditionally, astronomers estimate black hole mass using indirect measurements, such as velocity dispersion, luminosity, and other spectral properties. Machine learning can automate this process by finding complex relationships between these features and known black hole masses, making predictions more efficient and potentially more accurate.

2. Data Sources

For this project, we will use publicly available astrophysical datasets that contain black hole mass estimates along with other galaxy properties:

- **Sloan Digital Sky Survey (SDSS)** – Contains spectroscopic and photometric data for millions of galaxies.
- **NASA/IPAC Extragalactic Database (NED)** – Provides information on galaxy dynamics and black hole masses.
- **Event Horizon Telescope (EHT) & Chandra X-ray Observatory** – X-ray and radio data related to black hole environments.
- **The Supermassive Black Hole Mass Database** – A curated list of black hole mass measurements from the literature.

Estimating Black Hole Mass Using Machine Learning

3. Features and Labels

To predict black hole mass (MBH), we can use various astrophysical properties as features:

- **Galaxy Properties:**
  - Stellar velocity dispersion (σ) – How fast stars move near the galactic center.
  - Bulge luminosity – The brightness of the central region of the galaxy.
  - Galaxy mass – Total mass of the host galaxy.
  - Effective radius – The size of the galactic bulge.
  - Metallicity – The abundance of heavy elements in the galaxy.
- **Emission & Spectroscopic Features:**
  - X-ray and radio emissions – High-energy signatures associated with black holes.
  - Broad-line region width – Width of emission lines from fast-moving gas around the black hole.
  - Eddington ratio – The ratio of observed luminosity to the theoretical Eddington limit.
- **Target Variable (Label):**
  - Logarithm of black hole mass (logMBH), typically measured in solar masses.

4. Machine Learning Approaches

(still undecided which approaches will be implemented)

Several ML techniques can be used for regression-based mass estimation:

1. **Traditional Regression Models**
   - Linear Regression
   - Random Forest Regressor
   - XGBoost (Extreme Gradient Boosting)
2. **Deep Learning Approaches**
   - Multilayer Perceptron (MLP) – Neural networks trained on tabular astrophysical data.
   - Graph Neural Networks (GNN) – If we include relationships between black holes and their host galaxies.
3. **Hybrid Approaches**
   - **Physics-Informed Neural Networks (PINNs)** – Neural networks that integrate astrophysical laws into training.
   - **Gaussian Process Regression (GPR)** – Useful for capturing uncertainties in black hole mass predictions.

Estimating Black Hole Mass Using Machine Learning

5. Implementation Plan

1. **Data Collection & Preprocessing**
   - Download and clean datasets (e.g., remove missing values, normalize features).
   - Feature engineering (log transformations, standardization).
   - Data augmentation (e.g., generating synthetic data from known distributions).
2. **Exploratory Data Analysis (EDA)**
   - Visualize feature correlations (e.g., velocity dispersion vs. black hole mass).
   - Plot histograms and distributions of galaxy properties.
3. **Model Training & Validation**
   - Split dataset into training and test sets.
   - Train different ML models and compare performances.
   - Use cross-validation and hyperparameter tuning (e.g., GridSearchCV for Random Forest).
4. **Evaluation Metrics**
   - **Mean Absolute Error (MAE)** – Measures absolute differences in predictions.
   - **Root Mean Squared Error (RMSE)** – Penalizes larger prediction errors.
   - **R2 Score** – Indicates how well the model explains variance in the data.
5. **Interpretability & Results Analysis**
   - Feature importance analysis (e.g., SHAP values).
   - Compare ML-predicted masses to observed values in astrophysical studies.
6. **Deployment & Visualization**
   - Deploy an interactive dashboard using Streamlit or Flask API.
   - Allow users to input galaxy properties and get mass predictions.

Estimating Black Hole Mass Using Machine Learning

6. Expected Outcomes

- A trained ML model that can estimate black hole masses with high accuracy.
- Identification of the most important astrophysical features contributing to mass estimation.
- A deeper understanding of how machine learning can assist in astronomy.