|  |  |
| --- | --- |
| Cheat Sheet: Linear Regression | |
| Aim: Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables.  Base Model approach: DATA = TRUE SIGNAL + NOISE  Vanilla regression model:    .  Attributes: Homoscedasticity and independence of errors  Linear Basis Function Model+…+  Finding weights for ω0 to ωj   1. Use Log Likelihood function for Gaussian distribution: 2. Derive 1. to get the score function in respect to the gradient . 3. Results in: 4. Introduce design matrix for different 5. This finally results in:   Meaning: To calculate the weights, simply multiply the pseudo-inverse with the vector of targets.  Application: As a rule of thumb, use the pseudo-inverse only for less than 10000 samples.  Bias Variance Decomposition  Aim: The error of the model should be decomposed into an error that arises from a mismatch between the model and the real data **(bias)** and an error that arises from the noise in the data **(variance)**.   1. Use the expected value of the squared error between true values and predictions **(L2 error function)**:      1. Expanding and reforming yields the expected value of the noise and the expected value of the squared error between the real function f and the predictions:      1. Further expanding and reforming results in:     So, the general expected value of the error of a model depends on the noise of the data, the squared bias and the variance of the model across different datasets.  Analysis: ↑ model complexity: ↓ bias error term, ↑ variance error term    Good fit: The model is exactly is exactly as complex as it needs to be.  Overfit: The model is much too complex for the data.  Underfit: The model is not complex enough.  Regularization  Aim: Prevent the model from overfitting without reducing its complexity.  Idea: Keep the weights of the model small, as large weights cause a high sensitivity.  Therefore, the error function is defined as follows:    with the error E\_D between true values and predictions, the **regularization coefficient Ew** and the **regularization parameter** **λ**.    L1 norm:    **→ Lasso regularization** | L2 norm:  **→ Quadratic regularization**  The regularization becomes stronger with increasing λ as this leads to smaller model weights.  L1 regularization results in a rather sparse weight vector since some weights are set to zero.  L2 regularization primarily prevents the weights from becoming too large (due to the squaring).  **Basis Functions – Jeweils mit Graphenverlauf, wenn genügend Platz da ist?**  Polynomial Basis Function:  Gaussian basis functions  Sigmoidal basis functions  Periodic basis function  Bin-based basis function |