Regularization (a few notes)

Alipio Jorge (DCC-FCUP)

December 2020

Learning as search

- Initial model
  - random MLP
  - one node Decision Tree
- **Improve the model** with respect to data
  - reduce the **loss**
  - $Loss = f(Y, \hat{Y})$
  - the main part of loss is the **error**

Reduce search

- Search space can be large
- Simpler models are more likely than more complex models
  - Ockam's razor
- **Overfitting**
  - Simpler models are less prone to overfitting

Avoid complex models

- Many techniques to keep models simple
  - strong assumptions
    - linear regression
  - **penalize** more complex models
    - **complexity** parameter in Decision Trees

Regularization

- Parameter **shrinking**
  - add a **term** to the Loss function
  - this term penalizes **large** parameter values θ
  - it is called a **regularization** term

$$Loss(Y, \hat{Y}, \theta) = Error(Y, \hat{Y}) + \lambda . f(\|\theta\|)$$

- λ is the regularization **hyper-parameter**
  - if $\lambda = 0$ there is no regularization

Regularization

- Neural Networks
  - learning is done by **minimizing** the error
  - **Backpropagation** algorithm
    - usually admits a regularization term
    - to keep $\|W\|$ small
  - MLP for regression (α is the reg. hyper parameter)

$$Loss(y, \hat{y}, \theta) = \frac{1}{2}\|y - \hat{y}\|_2^2 + \frac{\alpha}{2}\|W\|_2^2$$

Regularization

- SVM
  - SVC implementation has two regularization hyper-parameters
    - `C` and `shrinking`
  - The function to **minimize** is

$$C \sum_i Loss(y_i, \hat{y}_i) + \Omega(w)$$

- Notes
  - the i loss corresponds to ξ_i
  - the w are the parameters of the separating hyper-plane

Hyperparameters C and Ω

in SVC

- `C`
  - is 1 by default
  - is cost factor
  - large `C` penalizes errors and keeps margins small
    - uses **fewer** support vectors
  - small `C` allows wider margins
- `shrinking`
  - `True` by default
  - defines if a **shrinking heuristic** is used
    - the algorithm decides if w are shrinked or not
  - **may** be useful to speed up computation

References

- Books
  - Han, Kamber & Pei, Data Mining Concepts and Techniques, Morgan Kaufman.
  - Jake VanderPlas, Data Science Handbook, O'Reilly
  - Tibshirani et al., Elements of Statistical Learning
- Scikit docs
  - https://scikit-learn.org/stable/modules/neural_networks_supervised.html#regularization
  - https://scikit-learn.org/stable/modules/svm.html#shrinking-svm