

## Machine Learning

### Chapter 4 - Anatomy of a Learning Algorithm

#### A learning algorithm

- A loss function
- An optimization criterion
- A way to find solutions to the optimization criterion

#### Gradient descent / stochastic gradient descent

- When the optimization is differentiable
- E.g. to optimize the linear regression function = to minimize the mean squared error

$$l \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N (y_i - (wx_i + b))^2.$$

- Take partial derivative of w and b
  - See going which side will finally reach a local minimum
- Sensitive to learning rate, slow for large dataset => New algos are proposed.

#### Mini-batch stochastic gradient descent

- Use smaller batches (subsets)
- Upgrades
  - Adagrad - scales a for each parameter according to gradient history.
  - Momentum - orients the gradient descent in the relevant direction and reducing oscillations.
  - RMSprop, Adam for NNWs

#### Features of different algorithms

- Numerical features only
  - SVM, logistic regression, linear regression, KNN, decision tree, etc..
- Categorical
  - Decision tree
- Weight => weight influences how decision boundary is drawn. It tells the machine which features are more important
  - SVM, logistic regression...
- Used both for classification and regression
  - SVM, KNN, decision tree...
  - Others only one type