ML Week 0x04 Logistic Regression

linear vs logistic ($\times 2$)

- y is discrete: classification
- Examples:
  - spam/non-spam
  - transaction: fraud or legitimate
  - tumor: malignant or benign
- So 0 or 1
- Problems with linear regression here (picture)
- sigmoid ($\times 2$)

Non-linear decision boundaries ($\times 2$)

- Still just use gradient descent *(algorithme du gradient) (de la plus forte pente) (de la plus profonde descente) (hill climbing)*
- This is why we like things to be differentiable
- Multinomial (multi-class) classification
  - one vs all (OvO, OvR) (draw picture, get three classifiers)
    - At decision time, try $k-1$ classifiers, choose the one with the most $+1$ votes (highest probability)
    - Problem: learners see more negatives than positives
    - Problem: different confidence for difference decision boundaries
  - one vs one (OvO) (draw picture, get three classifiers)
    - At decision time, try $k(k-1)/2$ classifiers, choose the one with the most $+1$ votes (highest probability)

Cost function ($\times 7$)

- This is not convex
- So potentially many local minima

- Plot cost and explain what it means for $y \in \{0, 1\}$.
- Note that our convex cost function
  - is differentiable
  - can be derived from statistics using the principles of maximum likelihood estimation (maximum de vraisemblance)

Exercises

Error types

- Type I error =
- Type II error =
- True/false positive/negative

Performance metrics

- Accuracy
- Precision
- Recall
- F1 measure
- ROC (receiver operating characteristics = fonction d'efficacité du récepteur, courbe ROC) AUC (area under curve = l'aire sous la courbe)