ML Week 0xJ1-4 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 (OvA, 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

- Null hypothesis *(hypothèse nulle)*
- True/false positive/negative
- Type I error = incorrect rejection of null hypothesis (roughly, false positives)
- Type II error = failure to reject null hypothesis (roughly, false negatives)
- 100% sensitivity = no false negatives
- 100% specificity = no false positives

Performance metrics

- Precision (also: sensitivity)
- Recall
- Accuracy
- F1 score (F1 measure) *roughly, a weighted average*
- *skip this:*
  - ROC (receiver operating characteristics = fonction d'efficacité du récepteur, courbe ROC) AUC (area under curve = l'aire sous la courbe)
  - Inventé pendant la WWII pour montrer la séparation entre les signaux radar et le bruit de fond.
- Confusion matrix *(matrice de confusion)* or contingency table, error matrix. Rows are classes, columns are predicted classes.

ROC

- x-axis is FPR, y-axis is TPR
- Plots cdf of TPR against cdf of FPR
- Really, plotting both against some classifier parameter
- Random guessing gives point on diagonal line. Coin flipping evolves towards $(.5, .5)$.
- Want: just an upper left point