ML Week   0x03   Logistic Regression

linear vs logistic (×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 (×2)

Non-linear decision boundaries (×2)

- Still just use gradient descent

- 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 (×7)

- This is not convex

- So potentially many local minima

- $\boxed{\text{Plot cost}}$ 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.