Boosting Charts

Karl Winterling

July 19, 2016

High Level

Let H_{0i} be the event that tissue sample $\mathbf{x}_i$ is cancerous.

- **Input:** weak learner algorithm L (classification error $\epsilon < 1/2$), preprocessed learning set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ where $\mathbf{x}_i$ is a vector of features representing a sample and $y_i = 1$ for "yes" (H_{0i} is true) and $y_i = -1$ for "no" (H_0 is false), number of "boosts" N.
- **Output:** Strong classifier C.
- Let D_1 define a probability distribution so that D_{1i} is the probability of choosing $\mathbf{x}_i$ in a sample.
- Pick a test set
- Do the following N times
  - Use D_t to sample the $\mathbf{x}_i$ with replacement to produce a learning set $S_t = \{(\mathbf{x}_i, y_i)\}$
  - Train L on S_t to produce a classifier C_t.
  - Determine the error of C_t on the entire set by comparing each $C_t(\mathbf{x}_i)$ to y_i and weighting by D_t.
  - Get α_t, where $\alpha_t = 0$ means C_t is a fair coin flip and higher α_t means C_t is better.
  - Use the error to weight C_t and add it to the strong classifier C.

  - Update D_t to D_{t+1} to so that S_{t+1} has more points C_t classified incorrectly.

- Return C

Low Level

- Determine the error ϵ_t of C_t

  - Let
$$\epsilon_t = \sum_{C_t(\mathbf{x}_i) \neq y_i} D_{ti}$$
    Define a "convenient" value,
$$\alpha_t = \frac{1}{2} \log \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) = \log \left(\sqrt{\epsilon_t{}^{-1} - 1} \right)$$
    A bigger α_t means ϵ_t is smaller. Think of α_t as a measure of how much C_t knows

- Update to $D_{t+1,i} = Z D_{ti} e^{-\alpha_t}$ if $C_t(\mathbf{x}_i) = y_i$ or $D_{t+1,i} = Z D_{ti} e^{\alpha_t}$ (don't hard code this!).

  - C_{t+1} needs to work on what C_t got wrong.
  - Makes sense if $\epsilon_t < 1/2$ (the weak learner L is better than classifying by flipping a fair coin), so $\alpha_t > 0$.

- Use C to classify $\mathbf{x}$ by,
$$C(\mathbf{x}) = \text{sign} \left(\sum_{t=1}^{N} \alpha_t C_t(\mathbf{x}) \right)$$

  - We give more weight to a learner C_t if ϵ_t is smaller (and therefore α_t is larger, but we're careful not to make α_t too big).
  - If a weak learner is similar to a coin flip, α_t is close to 0.