Week 2 Quiz

Spam Filters

We want to construct a spam filter classifier using logistic regression (or perceptron). The input data consists of a list of (email,spam/not spam label), and each email is represented by a variable length text string. Can we do that? What issues do you see and do you have any ideas how they could be adressed.

Generalization (Forgotten Last Time)

We want to learn $f : X \to \{-1, +1\}$. Assume f is probabilistic and independent of the input domain X. $P(f(x) = 1 \mid x) = P(f(x) = 1) = 0.2$. What this means is that when f is evaluated it ignores the input and flips its private biased coin independently of all other invocations and returns the result. What is the best out of sample error we can achieve? What if $P(f(x) = 1) = 0.5$? What are the optimal classifiers for these two cases?

Out Of Sample Error for Linear Regression with Noise

Assume the target function (probability distribution) is

$$p(y \mid x) = w^\intercal x + \varepsilon,$$

for some unknown w where ε is a stochastic noise term independent of x. Assume furthermore that ε is a zero mean Gaussian with standard deviation 1, meaning that expectation of the noise term ε squared is one i.e. $\mathbb{E}[\varepsilon^2] = 1$. In other words, given x, the target y is distributed as a standard Gaussian around $w^\intercal x$.

What is the best *out of sample error* possible for this target function using the least squares error function, $e(x, y) = (x - y)^2$?

Hint: What is the optimal classifier? What is the out of sample error of that one?

Break Points and Growth Functions

- Is there always a break point for a finite hypothesis set of n hypotheses? If so, can you give a an upper bound? What is the growth function?
- Does the set of all functions have a break point? What is its growth function?

- What is the (smallest) break point for the hypothesis set consisting of circles centered around $(0,0)$? For a given circle the hypothesis returns 1 for points inside the circle and -1 for points outside. What is the growth function?
- What if we move to balls in the 3-dimensional space $\mathbb{R}^3$? Or in general d-dimensional space $\mathbb{R}^d$ (hyperspheres)?
- Does the Nearest Neighbour hypothesis set have a break point. The Nearest Neighbour classifier returns the label of the nearest point in the training data (for example in euclidian distance). Hint: One may think of the hypothesis set as the set of all Voronoi Diagrams.

Convex Functions

Which of the following functions are convex?

- $f(x) = 2$
- $f(x) = -\ln(x), x > 0$
- $f(x) = x^3$
- $f(x) = x^2 + x^4$

Problem 1.12 in the book I will rewrite the problem a little bit. The input is n numbers $y_1, \ldots, y_n$. Define the minimization problem of finding the real number h that is on average closest to the n inputs measured by squared distance (least squares).

$$\arg\min_h \sum_{i=1}^{n} (h - y_i)^2$$

show that $h = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the mean value which we denote h_{mean}.

Show that if we change the the distance to absolute deviation i.e. compute

$$\arg\min_h \sum_{i=1}^{n} |h - y_i|$$

then $h = \text{median}(y_1, \ldots, y_n)$ is the median of the elements which we denote h_{med}.

Finally what happens to the solutions $h_{\text{mean}}, h_{\text{med}}$ if we add noise the last element y_n, i.e. $y_n = y_n + \varepsilon$ for $\varepsilon \to \infty$.