CPSC 531H Machine Learning Theory (Term 2, 2013-14)
Assignment 3

Due: Tuesday March 18th, in class.

Question 1: Can Johnson-Lindenstrauss Lemma preserve area?

(a): Suppose the distances between three points are preserved with multiplicative error $1+\epsilon$. Is the area of the corresponding triangle also always preserved with multiplicative error $1+O(\epsilon)$, or even some constant multiplicative error?

(b): Suppose u and v are mutually orthogonal unit vectors. Observe that the vectors u and v together with the origin form a right-angled isosceles triangle with area 1/2. Suppose the lengths of the triangle are distorted with multiplicative error at most $1+\epsilon$. What is the multiplicative error for the area of the triangle?

(c): Suppose a set V of n points are given in Euclidean space R^n. Let $0<\epsilon<1$. Give a randomized algorithm that produces a low-dimensional mapping $f: V \to R^T$ such that the areas of all triangles formed from the n points are preserved with multiplicative error $1+\epsilon$. What is the value of T for your mapping? Please give the exact number and do not use big-O notation.

Hint: If two triangles lie in the same plane (a 2-dimensional affine space) in R^n, then under a linear mapping their areas have the same multiplicative error. For every triangle, add an extra point to form a right-angled isosceles triangle in the same plane.)

Question 2: Margin Preservation under Dimensionality Reduction

Let $(x_1, y_1), \ldots, (x_N, y_N)$ be labeled examples with $x_i \in \mathbb{R}^n$, $\|x_i\| = 1$ and $y_i \in \{-1, 1\}$ for all i. Suppose that the examples are separable with margin γ. That is, there exists a vector $w \in \mathbb{R}^n$ with $\|w\| = 1$, $y_i = \operatorname{sign}(w^\mathsf{T} x_i)$ for all i, and $\min_i |w^\mathsf{T} x_i| = \gamma$.

Let $f: \mathbb{R}^n \to \mathbb{R}^k$ be the linear map $f(x) = Lx$ where L is the JL matrix of size $k \times n$. and $k = \Omega(\log(N)/\alpha^2\gamma^2)$. Prove that the dimension-reduced examples $(f(x_1), y_1), \ldots, (f(x_N), y_N)$ are separable with margin $(1-\alpha)\gamma$.

Hint: JL says that f preserves norms up to a factor $(1 \pm \epsilon)$ with high probability where, say, $\epsilon = \alpha\gamma/100$. Use the identities

$$\|u+v\|^2 - \|u-v\|^2 = 4u^\mathsf{T} v$$
$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2,$$

show that inner products are preserved with additive error. That is,

$$|f(u)^\mathsf{T} f(v) - u^\mathsf{T} v| \;\leq\; O(\epsilon).$$

Conclude that $f(w)$ is a consistent linear separator of the dimension-reduced examples, and

$$\min_i \frac{|f(w)^\mathsf{T} f(x_i)|}{\|f(w)\|\,\|f(x_i)\|} \;\geq\; (1-\alpha)\gamma.$$

Question 3: Subspace Embeddings: Recursive Version

In this question we gave a crude argument that gives a subspace embedding with $m = O(k \log(k/\alpha)/\alpha^2)$. In this question we will improve it to $m = O(k \log(1/\alpha)/\alpha^2)$.

As in class, let B be an $m \times k$ matrix whose entries are independent with distribution $N(0, 1/m)$. Let $S = \{\, x \in \mathbb{R}^k \,:\, \|x\|_2 = 1 \,\}$. Our goal is to show that

$$1 - \alpha \leq \|By\| \leq 1 + \alpha \qquad \forall y \in S.$$

Let P be an ϵ-net of S. Consider any point $y_0 \in S$. As before we approximate y_0 using $p_0 = \operatorname{argmin}_{p \in P} \|y_0 - p\|$. The residual is $y_1 = y_0 - p_0$ where $\|y_1\| \leq \epsilon$. As in class, we write $\|By_0\| \leq \|Bp_0\| + \|By_1\|$, and we use JL to control $\|Bp_0\|$.

The new idea is as follows. Instead of using the crude entry-wise bound on B to control the error $\|By_1\|$, you can also use P to approximate y_1. This will lead to additional vectors p_2 and y_2, then p_3 and y_3, etc.

Show that this strategy will work with $\epsilon = \alpha/3$ and $m = O(k \log(1/\alpha)/\alpha^2)$.

Question 4: BONUS: Improved Subspace Embedding

In this question we give an improved subspace embedding with $m = O(k/\alpha^2)$.

(a): Define $Q_\gamma = \left\{ w \,:\, w \in \frac{\gamma}{\sqrt{k}} \mathbb{Z}^k,\ \|w\|_2 \leq 1 \right\}$ for $\gamma \in (0,1)$. Prove $|Q_\gamma| \leq e^{k \cdot f(\gamma)}$ for some function $f(\gamma)$ (which needn't be optimized).

Hint: Given $z \in Q_\gamma$ define a cube C_z centered at z with side length $\gamma/\sqrt{k}$. Note that these cubes are all disjoint, then use a volume argument (you may use that an ℓ_2 ball of radius r in $\mathbb{R}^k$ has volume $(C_k \cdot r/\sqrt{k})^k$ for some constant C_k which is $\Theta(1)$ as k grows).

(b): Show that if for some matrix A of size $k \times k$ we have $|u^\mathsf{T} A v| \leq \alpha$ for all $u, v \in Q_\gamma$, then $|x^\mathsf{T} A x| \leq \alpha/(1-\gamma)^2$ for all $x \in \mathbb{R}^k$ of unit ℓ_2 norm.

Hint: Write $y = (1-\gamma)x$ and round down the coordinates of y to obtain $z \in Q_\gamma$. Argue that $y \in C_z$ and use that any point in a convex polytope can be written as a convex combination of the vertices of that polytope.

(c): Let B be the JL matrix of size $m \times k$ as before. Let A in the last problem be $B^\mathsf{T} B - I$. Choose some appropriate γ and conclude that $m = O(k/\alpha^2)$ suffices.