5 Appendix

5.1 Proof of Theorem 1

Proof. The weak learner assumption implies that for $x_k \in U^t$

$$\exists q \geq 0 : y_k \mathbf{h}(x_k)^T q > 0 \textit{ and } y_i \mathbf{h}(x_i)^T q > 0 \ \forall x_i \in L^t \tag{12}$$

Without loss of generality assume that $y_k = -1$. This implies that

$$\mathcal{A} = \left\{ q \geq 0,\, q \neq 0 \mid -\mathbf{h}(x_k)^T q > 0 \text{ and } y_i \mathbf{h}(x_i)^T q > 0 \ \forall x_i \in L^t \right\} \neq \emptyset \tag{13}$$

We are left to determine whether, there is a $q \geq 0$ such that, $\mathbf{h}(x_k)^T q > 0$ and $y_i \mathbf{h}(x_i)^T q > 0 \ \forall x_i \in L^t$. Suppose there is no such q, then we have that

$$\nexists q \geq 0 : \ \mathbf{h}(x_k)^T q > 0 \textit{ and } y_i \mathbf{h}(x_i)^T q > 0 \ \forall x \in L^t \tag{14}$$

By assumption $\mathcal{H}$ is negation complete that is $\exists j, j^* : h_j(x) = -h_{j*}(x)$. Define vector $\tilde{q}$ such that $\tilde{q}_j = q_j - q_{j*}$ then we can simplify the above expression to:

$$\nexists \tilde{q} : \ \mathbf{h}(x_k)^T \tilde{q} > 0 \textit{ and } y_i \mathbf{h}(x_i)^T \tilde{q} > 0 \ \forall x \in L^t \tag{15}$$

Note $\tilde{q}$ is now allowed to be negative. This means that as $\tilde{q}_i$ ranges over all the real numbers the vector $(\mathbf{h}(x_k)^T \tilde{q},\, y_1 \mathbf{h}(x_1)^T \tilde{q},\, \ldots,\, y_t \mathbf{h}(x_t)^T \tilde{q})$ does not intersect the first quadrant. In addition the complement of this set contains $\mathcal{A}$, which is convex and non-empty. Consequently, we can invoke the separating hyperplane theorem that separates the first quadrant from all the feasible vectors $(\mathbf{h}(x_k)^T \tilde{q},\, y_1 \mathbf{h}(x_1)^T \tilde{q},\, \ldots,\, y_t \mathbf{h}(x_t)^T \tilde{q})$ as $\tilde{q}_i, \forall i$ ranges over all real numbers. As a consequence we have hyperplane $\lambda \geq 0$ and $\delta > 0$ such that,

$$\exists \lambda,\, \delta \geq 0 : \delta \mathbf{h}(x_k)^T \tilde{q} + \sum_{i \in L^t} \lambda_i y_i \mathbf{h}(x_i)^T \tilde{q} \leq 0 \ \forall \tilde{q} \tag{16}$$

$$\exists \lambda,\, \delta \geq 0 : \ [\delta \mathbf{h}(x_k)^T + \sum_{i \in L^t} \lambda_i y_i \mathbf{h}(x_i)^T] \tilde{q} \leq 0 \ \forall \tilde{q} \tag{17}$$

$$\implies \delta \mathbf{h}(x_k) + \sum_{i \in L^t} \lambda_i y_i \mathbf{h}(x_i) = 0 \tag{18}$$

Note that λ or δ cannot be all zeros. For $\delta \neq 0$, equality in 18 implies that $\mathbf{h}(x_k)$ has to lie in the cone of $y_i \mathbf{h}(x_i)$'s. $\mathbf{h}(x)$ is a vertex of $+1, -1$ hypercube in N dimensions. A vertex $\mathbf{h}(x_k)$ of this hypercube lies in the cone of other vertices $\{\mathbf{h}(x_i)\}_{i \in L^t}$ if and only if $k \in L^t$.

For δ=0, the equality in 18 cannot hold for $\{y_i \mathbf{h}(x_i)\}_{i \in L^t}$ that satisfy the weak learner assumption.

□

5.2 Proof of Lemma 1

Proof. We provide the main outline of the proof and skip some of the messy algebra. For simpler notation, let $q(x) = \text{sgn}(\sum_{j=1}^{L} q_j h_j(x) - .5)$ where $h_j(x) \in \{0, 1\}$. We emphasize that the weak learners map to zero or one. Any two samples x, x' are δ-neighborly if:

$$\frac{1}{2} \int_Q |q(x) - q(x')| dq \leq \delta \tag{19}$$

The integral is the volume where $q(x)$ and $q(x')$ disagree:

$$\int_Q \mathbb{1}_{[q(x) \neq q(x')]} dq \leq 2\delta \tag{20}$$

Let $S = \{j | h_j(x) = h_j(x')\}$ and $S^c = \{j | h_j(x) \neq h_j(x')\}$:

$$q(x) = \text{sgn}(\sum_{j \in S} q_j h_j(x) + \sum_{j \in S^c} q_j h_j(x) - .5) \tag{21}$$

$$q(x') = \text{sgn}(\sum_{j \in S} q_j h_j(x) + \sum_{j \in S^c} q_j h_j(x') - .5) \tag{22}$$

Let $S_1 = \{j | h_j(x) = 1\} \cap S^c$ and $S_2 = \{j | h_j(x') = 1\} \cap S^c$ then

$$q(x) = \text{sgn}(\sum_{j \in S} q_j h_j(x) + \sum_{j \in S_1} q_j - .5) \tag{23}$$

$$q(x') = \text{sgn}(\sum_{j \in S} q_j h_j(x) + \sum_{j \in S_2} q_j - .5) \tag{24}$$

And $q(x) \neq q(x')$ if and only if

$$\sum_{j \in S} q_j h_j(x) < .5 \text{ and } \sum_{j \in S_1} q_j > .5 - \sum_{j \in S} q_j h_j(x) \text{ and } \sum_{j \in S_2} q_j < .5 - \sum_{j \in S} q_j h_j(x) \tag{25}$$

By the K-neighbor assumption: $|S_1 \cup S_2| \leq K$. Let $|S_1| = K - k_1$ and $|S_2| = k_1$ and:

$$\tilde{Q}(k_1) = \{q \in Q \mid \sum_{j \in S} q_j h_j(x) < .5, \ \sum_{j \in S_1} q_j > .5 - \sum_{j \in S} q_j h_j(x), \ \sum_{j \in S_2} q_j < .5 - \sum_{j \in S} q_j h_j(x)\} \tag{26}$$

It is easy to check that the case where $|S_2| = 0$ *and* $|S_1| = K$ will have the greatest volume:

$$Vol(\tilde{Q}(k_1)) \leq Vol(\tilde{Q}(0)) \ for \ 0 < k_1 \leq K \tag{27}$$

So let,

$$\tilde{Q}(0) = \{q \in Q, \sum_{j \in S_1} q_j > .5 - \sum_{j \in S} q_j h_j(x), \sum_{j \in S} q_j h_j(x) < .5\} \tag{28}$$

$Vol(\tilde{Q}(0))$ is an upper bound for (20).

To compute the volume we recast the problem in terms of probabilities. Note that since the simplex Q is endowed with the Lebesgue measure we can think of q as a random variable uniformly distributed over Q. However, the components of q are now dependent. To transform the problem into an independent set of random variables we consider exponentially distributed random variables.

Define the unnormalized IID random variable $q'_j = q_j \sum_{j \in 1}^{N} q'_j$ where q'_j are IID exponentially distributed random variables with mean equal to θ. Then $\mathbf{E}[\sum_{j \in 1}^{N} q'_j] = \frac{N}{\theta}$. It is well known that such an exponentially distributed set of random variables when normalized exactly produces a uniform distribution over the simplex.

By substitution of the unnormalized random variables we obtain,

$$\begin{aligned} Pr\{\tilde{Q}(0)\} =& Pr\{q \in Q, \sum_{j \in S_1} q_j > .5 - \sum_{j \in S} q_j h_j(x), \sum_{j \in S} q_j h_j(x) < .5\} \\ =& Pr\left\{ \sum_{j \in S_1} q'_j > .5(\sum_{j=1}^{N} q'_j) - \sum_{j \in S} q'_j h_j(x), \sum_{j \in S} q'_j h_j(x) < .5(\sum_{j=1}^{N} q'_j) \right\} \end{aligned}$$

To simplify this expression we consider the event,

$$A = \left\{ \left| \frac{1}{\theta} - \frac{1}{N} \sum_{j \in 1}^{N} q'_j \right| \leq \epsilon_2 \right\}$$

Note that the event A can be cast in the familiar form of an empirical average being close to its empirical mean. Consequently, we expect that the probability of the complement, A^c, of the event A is exponentially small in N.

We now proceed as follows:

$$\begin{aligned}
Pr\{\tilde{Q}(0)\} \leq & Pr\left\{\sum_{j\in S_1} q'_j > .5(\sum_{j=1}^{N} q'_j) - \sum_{j\in S} q'_j h_j(x), \sum_{j\in S} q'_j h_j(x) < .5(\sum_{j=1}^{N} q'_j),\ q'_j \in A\right\} + Pr(A^c) \\
& \leq Pr\left\{\sum_{j\in S_1} q'_j > .5\frac{N}{\theta}(1-\epsilon_2) - \sum_{j\in S} q'_j h_j(x), \sum_{j\in S} q'_j h_j(x) < .5\frac{N}{\theta}(1+\epsilon_2),\ q'_j \in A\right\} + Pr(A^c) \\
& \leq Pr\left\{\sum_{j\in S_1} q'_j > .5\frac{N}{\theta}(1-\epsilon_2) - \sum_{j\in S} q'_j h_j(x), \sum_{j\in S} q'_j h_j(x) < .5\frac{N}{\theta}(1+\epsilon_2)\right\} + Pr(A^c) \qquad (29)
\end{aligned}$$

where the first inequality follows from the union bound; the second inequality follows from the definition of event A; the third inequality is a direct application of the union bound. We now ignore the second term since it is arbitrarily small for sufficiently large N.

We are now in the familiar territory of a sum of IID random variables since S and S_1 have no overlap. Note that $\sum_{j\in S_1} q'_j$ is independent of $\sum_{j\in S} q'_j h_j(x)$ and each of these random variables are Γ distributed. By straighforward conditioning on $\sum_{j\in S} q'_j h_j(x)$ we can simplify the expressions in Equation 29. It follows that,

$$Pr\{\tilde{Q}(0)\} \leq \int_0^{.5} Pr\{\sum_{j\in S_1} q'_j > g\frac{N}{\theta}\}dg \qquad (30)$$

Let $Z = \sum_{j\in S_1} q'_j$ which has a gamma distribution: $\Gamma(K,\theta)$ and by the Chernoff bound(Section 5.2.1),

$$\begin{aligned}
Pr\{Z > g\frac{N}{\theta}\} &\leq \min_{t\geq 0} e^{-tg\frac{N}{\theta}}\mathbf{E}[e^{tZ}] \\
&= \min_{t\geq 0} e^{-tg\frac{N}{\theta}}(1-\frac{t}{\theta})^{-K}, t<\theta \\
&= (\frac{N}{K})^K e^K g^K e^{-gN}, g > \frac{K}{N}
\end{aligned}$$

The integral in (30):

$$= \int_0^{\frac{K}{N}} Pr\{\sum_{j\in S_1} q'_j > g\frac{N}{\theta}\}dg + \int_{\frac{K}{N}}^{.5} (\frac{N}{K})^K e^K g^K e^{-gN} dg \qquad (31)$$

The first term is upper-bounded by K/N since the integrand is positive and always less than 1. The second term is upper-bounded by:

$$\begin{aligned}
(\frac{N}{K})^K e^K \int_{\frac{K}{N}}^{.5} g^K e^{-gN} dg &\leq (\frac{N}{K})^K e^K \int_{\frac{K}{N}}^{\infty} g^K e^{-gN} dg \\
&= \frac{1}{N}\sum_{p=0}^{K} \frac{K!}{(K-p)!K^p} \\
&\leq \frac{K+1}{N}
\end{aligned}$$

Combining the bounds on the two terms, we have the upper bound:

$$Pr\{q(x) \neq q(x')\} \leq \frac{2K+1}{N} \qquad (32)$$

And the disagreement volume:

$$\int_Q \mathbb{1}_{[q(x)\neq q(x')]} dq \leq \frac{2K+1}{N} Vol(Q) \qquad (33)$$

And for any $Q' \subset Q$:

$$\int_{Q'} \mathbb{1}_{[q(x)\neq q(x')]} dq \leq \int_{Q} \mathbb{1}_{[q(x)\neq q(x')]} dq \leq \frac{2K+1}{N} Vol(Q) \tag{34}$$

$\square$

5.2.1 Chernoff Bound on a Gamma distribution

$$Pr\{Z > g\frac{N}{\theta}\} \leq \min_{t\geq 0} e^{-tg\frac{N}{\theta}} \mathbf{E}[e^{tZ}] \tag{35}$$

For a Gamma Random Variable $Z \sim \Gamma(K, \theta)$ the moment generating function is

$$\mathbf{E}[e^{tZ}] = (1 - \frac{t}{\theta})^{-K},\ if\ t < \theta \tag{36}$$

Minimize the bound over $0 \leq t < \theta$:

$$\mathcal{B}(t) = \frac{1}{e^{tg\frac{N}{\theta}}(1-\frac{t}{\theta})^K} \tag{37}$$

Let $t = \gamma\theta$ and maximize $\mathcal{B}'^{-1}(\gamma)$ instead:

$$\gamma^* = argmax_{0\leq\gamma<1}\ e^{c\gamma N}(1-\gamma)^K \tag{38}$$

Take the derivative:

$$\frac{d\mathcal{B}'^{-1}}{d\gamma} = (1-\gamma)^{K-1} e^{gN\gamma}[-K + (1-\gamma)gN] \tag{39}$$

The derivative is zero only when the last product term is zero or:

$$\gamma^* = 1 - \frac{K}{gN} \tag{40}$$

Note since $K << N$, $\gamma^* < 1$ and if $c \geq \frac{K}{N}$ then $\gamma^* \geq 0$. Plugging γ^* back in:

$$\mathcal{B}'(\gamma^*) = (\frac{N}{K})^K e^K g^K e^{-gN},\ if\ g > \frac{K}{N} \tag{41}$$

5.2.2 Integral of the Chernoff Bound on a Gamma distribution

$$(\frac{N}{K})^K e^K \int_{g_0}^{\infty} g^K e^{-gN} dg = e^{-g_0 N} \sum_{p=0}^{K} g_0^{K-p} \frac{K!}{(K-p)! N^{p+1}} \tag{42}$$

Let $g_0 = \frac{K}{N}$,

$$= \frac{1}{N} \sum_{p=0}^{K} \frac{K!}{(K-p)! K^p} \tag{43}$$

Define a term in this series as $A_p = \frac{K!}{(K-p)!K^p}$ and calculate the ratio of two succeeding terms:

$$r = \frac{A_p}{A_{p+1}} = \frac{K}{K-p} \geq 1 \tag{44}$$

The series is decreasing and the first term $A_0 = 1$ thus

$$\sum_{p=0}^{K} A_p \leq K + 1 \tag{45}$$

And the integral is bounded:

$$(\frac{N}{K})^K e^K \int_{g_0}^{\infty} g^K e^{-gN} dg \leq \frac{K+1}{N}, g_0 = \frac{K}{N} \tag{46}$$

5.3 Proof of Lemma 2

The proof closely follows [Nowak, 2009].

Proof. $\exists\ p'$ such that

$$|\sum_{i=1}^{B} q(x_i)p_i'| \leq \rho\ \forall q \tag{47}$$

Integrate both sides over $q \in Q'$

$$\int_{Q'} |\sum_{i=1}^{B} q(x_i)p_i'|dq \leq \rho\ Vol(Q') \tag{48}$$

Integral of the absolute value is greater than the absolute value of the integral and interchange integration with addition:

$$|\sum_{i=1}^{B} \int_{Q'} q(x_i)dq\ p_i'| \leq \rho\ Vol(Q') \tag{49}$$

If $x \in \mathcal{X}$ s.t. $|\int_{Q'} q(x)dq\ | \leq \rho\ Vol(Q')|$ does not exist then $|\int_{Q'} q(x)dq\ | > \rho\ Vol(Q')$ for all $x \in \mathcal{X}$. Since (49) is a convex combination of $\int_{Q'} q(x_i)dq$, if one term is negative there has to exist a positive term in order for the sum to be less than or equal to $\rho\ Vol(Q')$. Therefore $\exists\ x,\ x'$ such that:

$$\int_{Q'} q(x)dq > \rho\ Vol(Q')\ and\ \int_{Q'} q(x')dq < -\rho\ Vol(Q') \tag{50}$$

If the pair $Q, \mathcal{X}$ is δ-neighborly, there exists a sequence of x_i's starting at x and ending in x'. The sign will have to switch somewhere in the sequence. Let us redefine the pair x, x' to be where the sign switches. From before: $\int_{Q'} q(x)dq - \int_{Q'} q(x')dq > 2\rho\ Vol(Q')$. By δ-neighborly assumption: $|\int_{Q'} q(x)dq - \int_{Q'} q(x')dq| < \int_{Q'} |q(x) - q(x')|dq < 2\delta Vol(Q)$. Combining the two inequalities: $Vol(Q') < \frac{\delta}{\rho} Vol(Q)$.

□

5.4 Proof of Theorem 2

Proof. Let $\rho \geq \rho^*\{\mathcal{X}, Q\}$ and at this stage we want to find an x' to reduce version space Q^τ by $\frac{1+\rho}{2}$ at stage τ. Lemma 2 states that if that is not possible then

$$Vol(Q^\tau) \leq \frac{\delta}{\rho} Vol(Q) \tag{51}$$

For simplicity of notation call this the termination of stage 1 and let τ be the time stage 1 is terminated, namely, the condition above is realized.

To proceed we now restart the entire process by exchanging Q with Q^τ. We call this start of stage 2. To avoid confusion we denote the iterations in this stage by t. Let $\rho_t \geq \rho^*\{\mathcal{X}, Q^t\}$. Observe that since $Q^t \subset Q$, $\rho^*(\mathcal{X}, Q^t) \leq \rho^*(\mathcal{X}, Q)$ and we can set $\rho^*\{\mathcal{X}, Q\} \leq \rho_t < 1$.

By following the proof of Lemma 2, at some time t if an x such that $|\int_{Q^t} q(x)dq\ | < \rho_t\ Vol(Q^t)$ does not exist than there must exist x and x' such that:

$$\int_{Q^t} q(x)dq - \int_{Q^t} q(x')dq > 2\rho_t\ Vol(Q^t) \tag{52}$$

Let $V_d(Q') = \int_{Q'} \mathbb{1}_{[q(x) \neq q(x')]} dq$. Let $Q_C^t = Q \setminus Q^t$ and $Vol(Q_C^t) \geq (1 - \frac{\delta}{\rho}) Vol(Q)$.

$$V_d(Q^t) + V_d(Q_C^t) = V_d(Q) \tag{53}$$

By the regularity assumption (9), $V_d(Q_C^t) \geq \alpha V_d(Q)$ and

$$V_d(Q^t) \leq (1-\alpha) V_d(Q) \tag{54}$$

And by δ-neighborly assumption, $V_d(Q) \leq \delta Vol(Q)$ and

$$V_d(Q^t) \leq (1-\alpha)\delta \; Vol(Q) \tag{55}$$

Combining this expression with inequality 52 we obtain:

$$Vol(Q^t) \leq \frac{(1-\alpha)\delta}{\rho_t} Vol(Q) \tag{56}$$

The first statement of Lemma 2 states that for any two consecutive version space Q^t and Q^{t+1} the following reduction is possible for $\rho^* \leq \rho < 1$ ($\rho^* := \rho^*\{\mathcal{X}, Q\}$)

$$Vol(Q^{t+1}) \leq \frac{(1+\rho)}{2} Vol(Q^t) \tag{57}$$

If this condition is not satisfied then the volume bound of Eq. 56 must hold. Now note that the ratio of the volume bound at the termination of the previous stage τ (see Eq. 51) and at the termination of the current stage t (see Eq. 56) is a constant equal to $(1-\alpha)$. Furthermore, we are guaranteed an exponential rate $(1+\rho_t)/2$ of decay while going from termination of stage 1 to termination of stage 2. Consequently, we can reduce the volume at the previous stage τ to the current stage t with at most a constant number of queries. For simplicity we assume that this is equal to one since the order-wise scaling of the number of queries does not change. Consequently, we can obtain:

$$Vol(Q^{t+1}) = \frac{(1-\alpha)\delta}{\rho} Vol(Q^t) \tag{58}$$

To obtain the worst case rate for each iteration we need:

$$\lambda_0 = \min_{\rho^* \leq c \leq 1} \max\{\frac{1+\rho}{2}, \frac{(1-\alpha)\delta}{\rho}\} \tag{59}$$

This expression simplifies to the situation when the two arguments are equal. This turns out to be $\rho = \frac{1}{2}(\sqrt{1+8(1-\alpha)\delta} - 1)$

$$\lambda_0 = \max\{\frac{1+\rho^*}{2}, \frac{1+.5(\sqrt{1+8(1-\alpha)\delta}-1)}{2}\} \tag{60}$$

where $\delta = \frac{2K+1}{N}$. We now note that $\sqrt{1+z} \leq 1 + z/2$. Consequently, we get,

$$\lambda_0 \leq \lambda = \max\{\frac{1+\rho^*}{2}, \frac{1}{2}(1+(1-\alpha)\frac{2K+1}{N})\}$$

We can repeat this argument for Stage 3, Stage 4 and so on in an identical fashion. The volume of our final version space is required to be $Vol(Q^n) = \epsilon Vol(Q)$.

$$Vol(Q^n) = \lambda^n Vol(Q)$$

$$\epsilon = \lambda^n \implies n = \frac{\log \epsilon}{\log \lambda}$$

□

5.5 Proof of Theorem 3

Proof. In the proof, all volume is taken with respect to the lebesgue measure on the p sparse subspace. If we can reduce the volume of sparse version space at each stage by λ then after n stages:

$$Vol(S^n) = \lambda^n Vol(S) \tag{61}$$

There are $\binom{N}{p}$ p-sparse disjoint segments: $\{s_1, s_2, \ldots, s_{\binom{N}{p}}\} = S$. Without loss of generality, we define the volume $Vol(.)$ such that $Vol(s_r) = 1$ for $r = 1, \ldots, \binom{N}{p}$ therefore

$$Vol(S^n) = \lambda^n \binom{N}{p}$$

By assumption from Section 3.2, we defne

$$q_s = \arg \inf_{q^* \in S} Vol\{q \in S \mid ||q - q^*||_1 \leq \frac{\theta}{2}\} \tag{62}$$

$$f(\theta, p) = Vol\{q \in S \mid ||q - q_s||_1 \leq \frac{\theta}{2}\} \tag{63}$$

If $Vol(S^n) \leq f(\theta, p)$ then $S^n \subset \{q \in S \mid ||q - q_s||_1 \leq \frac{\theta}{2}\}$ and $\forall q \in S^n$ (by the margin bound [Schapire et al., 1997])

$$Prob(q(x) \neq y) \leq O\left(\frac{\log |\mathcal{X}| \log p}{\theta^2 |\mathcal{X}|} + \frac{\log(1/\delta)}{|\mathcal{X}|}\right)^{\frac{1}{2}} \tag{64}$$

So we require:

$$\begin{aligned} Vol(S^n) &\leq f(\theta, p) && (65)\\ n \log \lambda + \log \binom{N}{p} &\leq \log f(\theta, p) && (66)\\ n &\geq \frac{\log \binom{N}{p} + \log \frac{1}{f(\theta,p)}}{\log \frac{1}{\lambda}} && (67) \end{aligned}$$

$\square$

5.6 Proof of Lemma 3

Proof. If $\rho^* < 1$ then $\nexists q \in Q$ s.t. $q^T \mathbf{h}(x_i) > 0 \ \forall i$. Let us define a vector $f(q) \in R^B$ with $f(q)_i = q^T \mathbf{h}(x_i)$ and a set $F = \{f(q) | q \in Q\}$. Since every component of f cannot be positive, the set F cannot lie in the first (positive) orthant. The set F is also convex, so there must exist a separating hyperplane with a normal vector $\lambda \geq 0$. This implies the following inequality:

$$\sum_{i=1}^{B} \lambda_i f(q)_i = \sum_{i=1}^{B} \lambda_i \sum_{j=1}^{N} q_j h_j(x_i) \leq 0 \tag{68}$$

At least one element of λ must be non-zero to define a hyperplane. Let us interchange the summation:

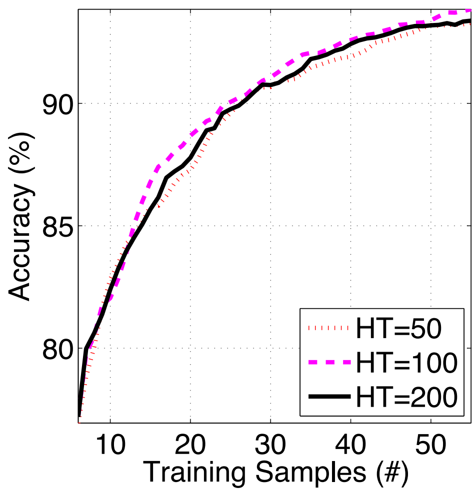
$$\sum_{j=1}^{N} q_j \sum_{i=1}^{B} \lambda_i h_j(x_i) \leq 0 \tag{69}$$

From earlier, we assume that for every weak hypothesis there exists a compliment: s.t. $h_j(x) = -h_{j^*}(x)$ and $h_j, h_{j^*} \in \mathcal{H}$. For any weight vector q, we can reassign the weight of h_j to its compliment h_{j^*} and make the left side in (69) greater than zero. But the inequality in (69) has to hold for all $q \in Q$. This can only be true if every term in the summation is zero:

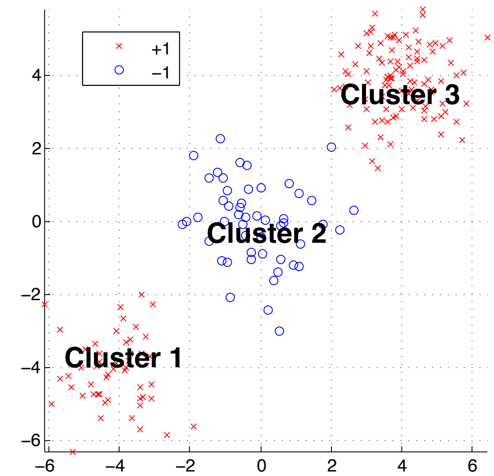
$$\sum_{i=1}^{M} \lambda_i h_j(x_i) = 0 \ \forall j \tag{70}$$

$\square$

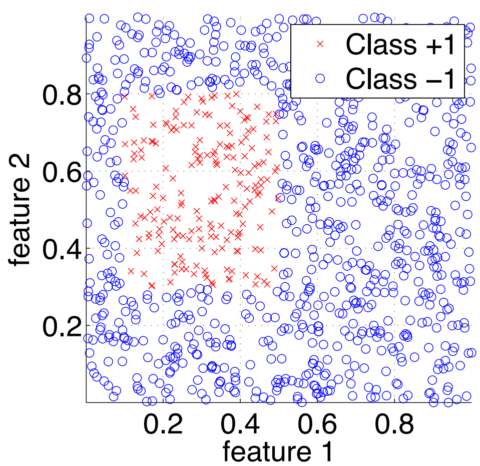
5.7 Miscellaneous Figures



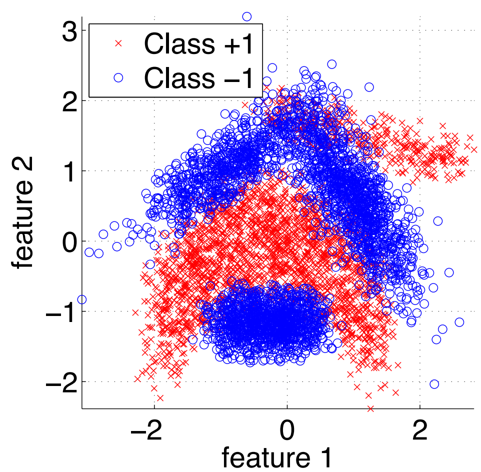
(a) Accuracy vs iterations of Hit and Run



(b) $2D$ Clusters



(c) Box Dataset



(d) Banana Dataset

Figure 7: Accuracy vs # labeled examples as a function of Hit and Run iterations (HT): changing HT does not change performance 7(a). Two dimensional dataset: Gaussian Clusters 7(b).Box Dataset 7(c). Banana Dataset 7(d).

5.8 Sampling with Hit and Run in the boosting framework

Algorithm 2 sample

INPUT: L^t {labeled set of examples}, T_s { number of iterations}, q^0 {initial feasible point}
$Q^t \leftarrow \{q : q \in Q, m_i^T q \geq v_0 \ \forall i | x_i \in L^t\}$, $Q \leftarrow \{q : q \geq 0, \mathbf{1}^T q = 1\}$, $d_0 \leftarrow \frac{1}{N}\mathbf{1} - q^0$ {initial direction} $w = \frac{1}{\sqrt{N}}\mathbf{1}$
for $s = 1$ to T_s **do**
  $z \leftarrow \mathcal{N}(0, I)$, $z' \leftarrow [I - ww^T]z$, $d \leftarrow \frac{z}{||z||_2}$
  {Generate a normal random variable, project it onto a hyperplane parallel to the simplex, and normalize to form a random direction }
  $r_i^1 \leftarrow \frac{(q_s)_i}{(-d)_i}$, $r_i^2 \leftarrow \frac{(M^t q_s - v_0)_i}{(-M^t d)_i}$
  $\alpha^+ \leftarrow min\{min_{r_i^1 \geq 0}\ r_i^1\ , min_{r_i^2 \geq 0}\ r_i^2\}$, $\alpha^- \leftarrow max\{max_{r_i^1 < 0}\ r_i^1\ , max_{r_i^2 < 0}\ r_i^2\}$
  $q_s^+ \leftarrow q_s + \alpha^+ d$, $q_s^- \leftarrow q_s + \alpha^- d$ {find two endpoints}
  $\alpha_s \leftarrow UNIFORM[0, 1]$ {generate a uniform random variable on $[0, 1]$}
  $q_{s+1} \leftarrow q_s^+ \alpha_s + q_s^-(1 - \alpha_s)$ compute new interior point
end for
OUTPUT: $q_{sample} \leftarrow q_{T_s}$ {uniform random sample from Q^t}