Alex Fay
HW #5

(1a) Prove $\|x_i - \sum_{j=1}^{k} z_{ij} v_j\|^2 = x_i^T x_i - \sum_{j=1}^{k} v_j^T x_i x_i^T v_j$

— Proof —

$$\|x - \sum_{j=1}^{k} z_{ij} v_j\|^2 = \left(x - \sum_{j=1}^{k} z_{ij} v_j\right)^T \left(x - \sum_{j=1}^{k} z_{ij} v_j\right)$$

$$= x^T x_i - 2 \sum_{j=1}^{k} z_{ij} v_j^T x_i + \sum_{j=1}^{k} v_j^T z_{ij}^T z_{ij} v_j$$

— $v^T v = 1$ when $i = j$ thus we can say the following —

$$= x^T x - 2 \sum z_{ij} v_j^T x_i + \sum v_j^T x_i x_i^T v_j$$

$$\boxed{= x^T x_i - \sum_{j=1} v_j x_i x_i^T v_j}$$

(1b) Prove $J_k = \frac{1}{n} \sum_{i=1}^{n} \left(x_j^T x_i - \sum_{j=1} v_j^T x_i x_i^T v_j\right) = \frac{1}{n} \sum_{i=1} x^T x_i - \sum_{j=1}^{k} \lambda_j$

— Proof —

Given $J_k = \frac{1}{n} \sum_{i=1}^{n} \left(x^T x_i - \sum_{j=1}^{k} v_j^T x_i x_i^T v_j\right)$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \frac{1}{n} \sum_{j=1}^{k} v_T \left(\sum_{i=1} x_i x_i^T v_j\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} v_j^T \Sigma v_j$$

$$\boxed{= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} \lambda_j}$$

1c

If $k = d$, there is no truncation so $J_d = 0$

Show error for $k < d$ is $\sum_{j=k+1}^{d} \lambda_j$

-Proof-

Given: $J_d = 0$

Thus, $\frac{1}{n}\sum_{i=1}^{n} x_i^T x_i = \sum_{j=1}^{d} \lambda_j$

Thus,

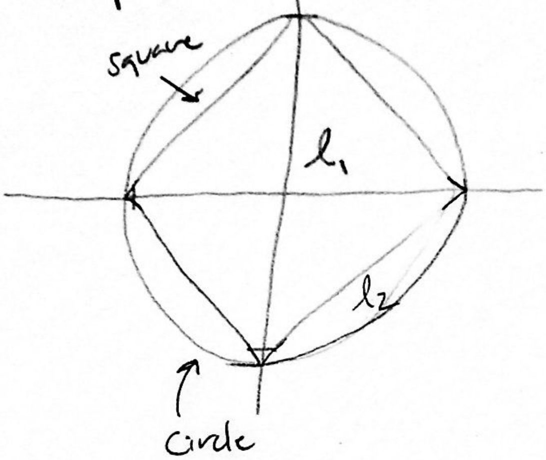
$$J_k = \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{d} \lambda_j + \sum_{j=k+1}^{d} \lambda_j$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i - \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i + \sum_{j=k+1}^{d} \lambda_j$$

$$= \sum_{j=k+1}^{d} \lambda_j$$

(2)

Graph for norm - ball



Show the optimization problems are equivalent

Since $-\lambda_k$ has no x_i then $\min f(x) + \lambda(\|x\|_p - k)$

$= \min f(x) + \lambda \|x\|_p$

Thus, opt $x = \min f(x) + \lambda \|x\|_p$ for $\lambda \geq 0$

Since l_2 is a circle + l_1 is a square w/ edges,
The probability of landing on an edge for $l_1 > l_2$.
This means that at upper/higher dimensions then
l_1 has more zero weights than l_2 (since face elements
are nonzero + edges are 0 weights).

Extra Credit

Show that placing an equal zero-mean balance on each element of weight $\theta = $ to l_1 on Max,

$\text{Max } P(\theta | D) = \text{Max } \log P(\theta | D)$

$P(\theta | D) = P(D | \theta) P\theta / P(D)$

Using log rules of division,

$\log P(\theta | D) = \log P(D|\theta) + \log P(\theta) - \log P(D)$

$-P(D)$ does not have θ part so -

$= \text{Min} -\log P(D|\theta) - \log P(\theta)$

-Given $\text{Lap}(x | \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right)$ then -

$-\log P(\theta) = \frac{1}{b} \sum \|\theta_i\| + z$

$= \lambda \|\theta\| + z$

From the above we know that $P(\theta | D) = \text{Min} -\log P(D|\theta) + \lambda \|\theta\|$

Graph Comparison of Laplace + Standard Norm Density: