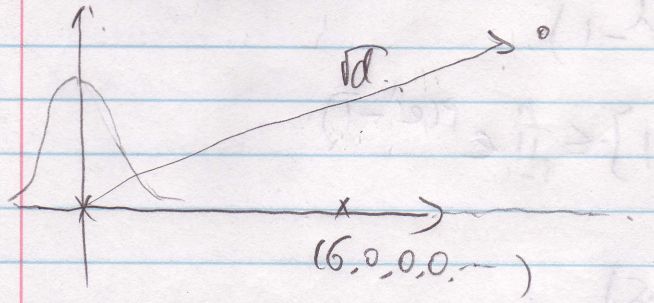
4/24

Scribe: Cheng Lu, Kaihan Yin

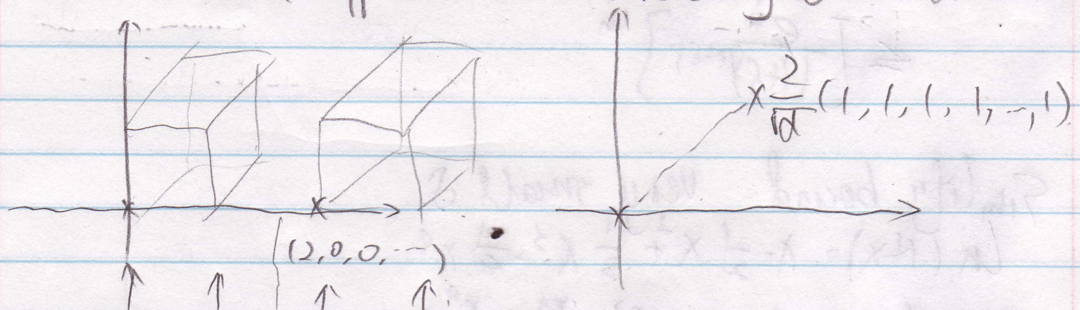
Apr. 24th

How close can centers be so that we can still distingish which center generetes a point

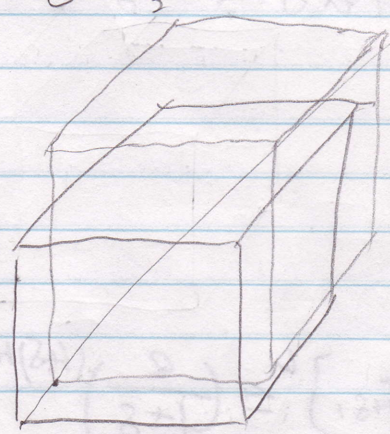
Gaussian

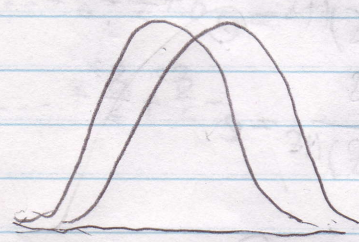


If distributions binomial cannot rotate coordinates
Thus it makes a difference as to where my centers are



$$X_i \begin{cases} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{cases}$$





Define radius R to be radius of sphere so sphere contains half of the probability mass.

$$R \leq \sqrt{2}\,\sigma \quad \text{if } \sigma \text{ exists}$$

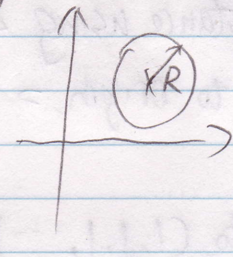
Variance can be arbitrarily large relative to R.

$\sigma^2 = \int x^2 p(x)\,dx \geq R^2 \cdot \frac{1}{2}$

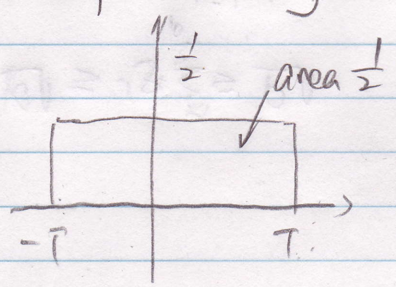
$R^2 \leq 2\sigma^2$

$R \leq \sqrt{2}\sigma$

Lemma: $R \leq \sqrt{2}\,\sigma$



Consider two processes with centers at the origin and $(1,1,1,\dots,1)$ and probability distribution



Each coordinate is independent and generated from given probability distribution

0 0 x x 0 x 0 x 0 x

x 1 x 1 x x 1 1 x x x

What if we classified a point by measuring distance to each center. What if we use the L_2 norm:

$x_1, x_2, x_3, \dots, x_{\frac{d}{2}}, 0, 0, \dots, 0$

$$d_0 = \sum_{i=1}^{\frac{d}{2}} x_i^2 = (\text{distance})^2$$

$$d_1 = (\text{distance to } (1,1,1,\dots,1))^2 = \sum_{i=1}^{\frac{d}{2}} (1-x_i)^2 + \frac{d}{2} = \sum_{i=1}^{\frac{d}{2}} (1-2x_i+x_i)^2 + \frac{d}{2}$$

$$= d - 2\sum_{i=1}^{\frac{d}{2}} x_i + \sum_{i=1}^{\frac{d}{2}} x_i^2$$

$d_1 - d_0 = d - 2\sum_{i=1}^{d} x_i$ $E(d_1 - d_0) = d$ $\because E(\sum_{i=1}^{d} x_i) = 0$

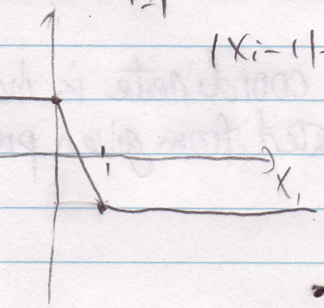
if $T \ge d$, will classify point randomly and misclassify half the time

Measure distance using L_1 norm

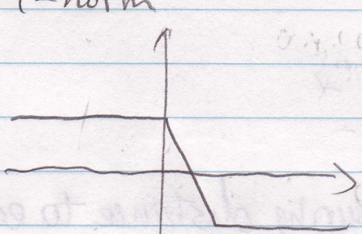
d_0 = dist to origin = $\sum_{i=1}^{\frac{d}{2}} |x_i|$

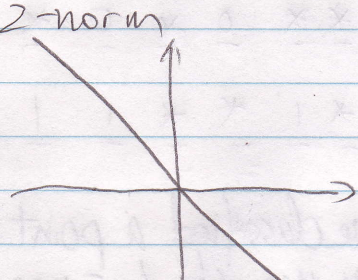
d_1 = dist to $(1,1,1,\cdots)$ = $\sum_{i=1}^{\frac{d}{2}} |x_i - 1| + \frac{d}{2}$

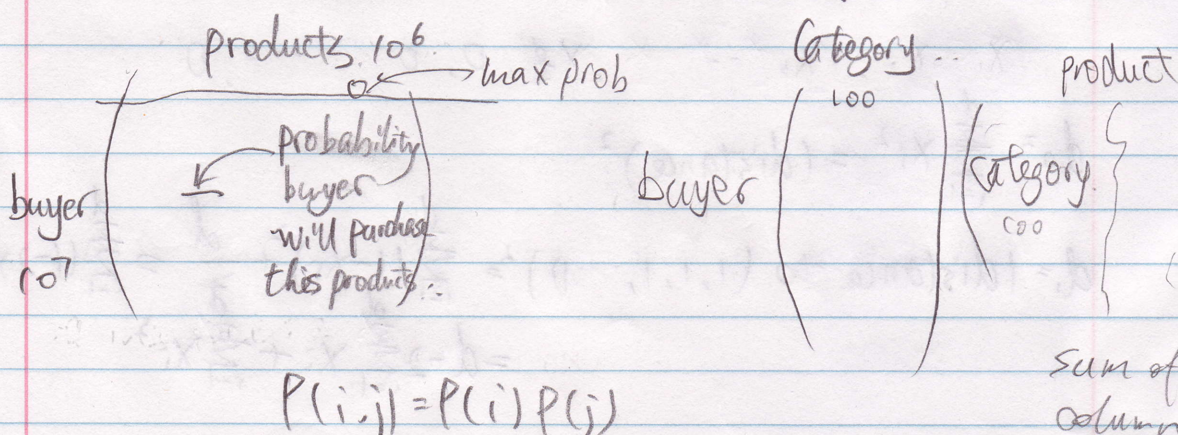
$$d_1 - d_0 = \frac{d}{2} + \sum_{i=1}^{\frac{d}{2}} (|x_i - 1| - |x_i|) = \frac{d}{2} + \sum_{i=1}^{\frac{d}{2}} \delta_i \qquad \delta_i = \pm 1$$



$$-\sqrt{d} \le \sum_{i=1}^{\frac{d}{2}} \delta_i \le \sqrt{d}$$







$P(i,j) = P(i)P(j)$

$P(i,j) = P(i)P(j)$

$a_1, a_2, \cdots, a_n$

$a_i = \{b_1, b_2, \cdots, b_k\}$

$E\left(h(b_j) \sum_{i=1}^{n} h(a_i) \right) = \#(b_j)$

	h_1	h_2	h_3	h_4	$\cdots$	h_8
b_1	-1	1	-1	1		
b_2	-1	-1	1	1		
b_3	-1	-1	-1	-1		

$2 \times (\mathrm{rand}(3,1) \leq \frac{1}{2}) - 1$

Variance

Kumar, Raghavan, Rajapolan, Tomkins.
Recommendation systems: a probabilistic analysis.

Jon Kleinberg and Mark Sandler
Using Mixture Models for Collaborative Filtering

$$\begin{array}{cc} & \text{items} \\ \text{buyer} & \begin{pmatrix} & A & \\ & & \end{pmatrix} \end{array} = \begin{array}{cc} & \text{categories} \\ \text{buyer} & \begin{pmatrix} P \end{pmatrix} \end{array} \begin{array}{c} \text{items} \\ \begin{pmatrix} W \end{pmatrix} \end{array}$$

prob

$m \times n$ $\qquad$ $m \times k$ $\qquad$ $k \times n$

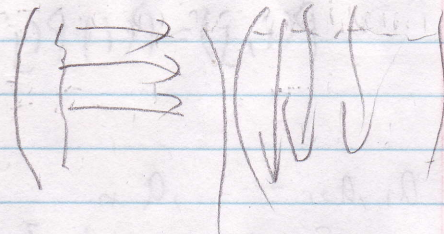
Rows of A, P and W sum to 1.

If I have a rank k matrix A whose rows sum to 1. Can I factor it into P and W whose rows sum to 1.

$A = (U\Sigma)U$

combine ↑ normalize rows → W

correct for row normization by normalizing columns → P



$$\sum_{l=1}^{n} A_{il} = 1$$

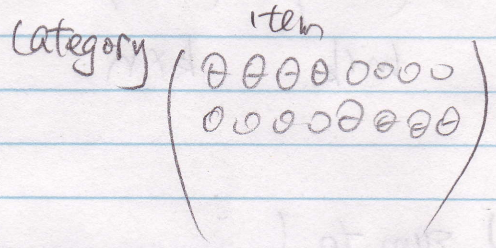
$$\sum_{l=1}^{n}\sum_{j}^{k} P_{ij} W_{jl} = 1$$

$$\sum_{j=1}^{k} P_{ij} \underbrace{\sum_{l=1}^{n} W_{jl}}_{1} = 1$$

$$\sum_{j=1}^{k} P_{ij} = 1$$

Two cases.
1). cluster disjoint
2) cluster can overlap

If we knew the categories and the categories were disjoint we could calculate W.



sample size = 1
sample size = 2

$\text{Prob}(i, j) \stackrel{?}{=} \text{Prob}(i)\,\text{Prob}(j)$

Correlation test for model with disjoint clusters

$\#(i)$ number of occurence of i

$\#(i,j)$ number of occurence of pair of items i and j

buyer $\left(\underline{\quad P \quad}\right)\left(\overset{\text{items}}{W}\right)$

$$\text{freq of } i = \sum_{u=1}^{W} P_{uC(i)} W_{C(i)i} = W_{C(i)i} \sum_{u=1}^{m} P_{uC(i)}$$

$$\text{freq of pair } i,j = \sum_{u=1}^{m} P_{uC(i)} W_{C(i)i} P_{uC(j)} W_{C(j)j}$$

$$= W_{C(i)i} W_{C(j)j} \sum_{u=1}^{m} P_{uC(i)} P_{uC(j)}$$

To determine if elements are in the same category considers two vectors

$$(P_{1C(i)}, P_{2C(i)}, \ldots, P_{mC(i)})\, W_{C(i)i}$$

$$(P_{1C(j)}, P_{2C(j)}, \ldots, P_{mC(j)})\, W_{C(j)j}$$

What if $C(i) = C(j)$
vectors are parallel

If clusters $C(i)$ and $C(j)$ are distinguishable vectors are not close to parallel

Test to determine if vectors are parallel

If $\#(i)$ and $\#(j)$ are large then $\#(i,j)$ is close to expected value of $\#(i,j)$.

$$E(\#(i,j)) = \sum_u P_{uC(i)} W_{C(i)i} P_{uC(j)} W_{C(j)j}$$

Consider $\dfrac{E(\#(i,j))\,E(\#(i,j))}{E(\#(i,i))\,E(\#(j,j))}$

$$\cancel{W_{c(i)} W_{c(j)}\ W_{c(i)} W_{c(j)}}\ \sum_u P_{uc(i)} P_{uc(j)} \sum P_{uc(i)} P_{uc(j)}$$

$$\cancel{W_{c(i)} W_{c(j)}\ W_{c(j)} W_{c(j)}}\ \sum_u P_{uc(i)} P_{uc(j)} \sum P_{uc(j)} P_{uc(j)}$$

$$x = (P_{1c(i)}, \cdots P_{mc(i)})$$

$$y = (P_{1c(j)}, \cdots P_{mc(j)})$$

$$= \frac{(xy)^2}{x^2 y^2} = \cos^2\theta$$