

Nicholas Miller

Math for Intelligent Systems

HW 2

MIS HW #2

2.8(a)

$$a^T c = \int_{-\beta}^{\beta} p(x) dx = c_1(x - \beta) + \frac{c_2}{2}(x^2 - \beta^2) + \dots + \frac{c_n}{n}(x^n - \beta^n)$$

$$\text{Therefore, } a = (x - \beta, \frac{x^2 - \beta^2}{2}, \dots, \frac{x^n - \beta^n}{n})$$

(b)

$$b^T c = p'(x) = c_2 + 2c_3x + 3c_4x^2 + \dots + (n-1)c_nx^{n-2}$$

$$\text{Therefore, } b = (0, 1, 2x, 3x^2, \dots, (n-1)x^{n-2})$$

2.12(a)  $\beta_3 < 0$  means that product three is negatively related to total profit, meaning an increase in price will result in loss of profit and a decrease in price will result in a gain of profit.

(b) The product I will choose will be the one with the largest absolute  $\beta$ . If the  $\beta$  was originally negative I will decrease the price by 1% and if the  $\beta$  was originally positive I would increase the price by 1%.

(c) For this part I would do the same as above but with the two largest absolute  $\beta$ s.

3.5 Lemmas; properties of absolute values.

i.  $|\beta a| = |\beta||a|$

ii.  $|x+y| \leq |x| + |y|$

iii.  $|x| \geq 0 \quad \frac{1}{2}|x|=0 \text{ iff } x=0$

1-norm

(1) homogeneity

$$\|\beta x\|_1 = \sum_{i=1}^n |\beta x_i| = \sum_{i=1}^n |\beta| |x_i| = |\beta| \sum_{i=1}^n |x_i| = |\beta| \|x\|_1$$

Therefore,  $\|\beta x\|_1 = |\beta| \|x\|_1$ .

(2) triangular inequality

$$\|x+y\|_1 = \sum_{i=1}^n |x_i+y_i| \leq \sum_{i=1}^n |x_i| + |y_i| = \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|x\|_1 + \|y\|_1$$

Therefore,  $\|x+y\|_1 \leq \|x\|_1 + \|y\|_1$ .

(3) definiteness

Assume  $\|x\|_1 = 0 \Leftrightarrow \sum_i |x_i| = 0 \Leftrightarrow \forall i |x_i| = 0$  (by (iii))  
 $\Leftrightarrow \forall i x_i = 0$  by (ii)  $\Leftrightarrow x = \vec{0}$

Therefore,  $\|x\|_1 = 0$  iff  $x = \vec{0}$

(4) nonnegativity

Assume  $0 = \|0\|_1 = \|x+(-x)\|_1 \stackrel{(2)}{\leq} \|x\|_1 + \|(-x)\|_1$   
 $\stackrel{(1)}{=} \|x\|_1 + 1 - \|x\|_1 = \|x\|_1 + 1 - \|x\|_1 = 1 \neq 0$

Therefore,  $0 \leq \|x\|_1$

(b) lemmas: properties of max

(iv) let  $x_j \neq 0$  then  $\|x\|_\infty = \max_i |x_i| \geq |x_j| > 0$

(1) homogeneity

$$\|\beta x\|_\infty = \max_i |\beta x_i| = \max_i |\beta| |x_i| = |\beta| \max_i |x_i| = |\beta| \|x\|_\infty$$

Therefore,  $\|\beta x\|_\infty = |\beta| \|x\|_\infty$

(2) triangular inequality

$$\begin{aligned} \|x+y\|_\infty &= \max_i |x_i + y_i| \leq \max_i (|x_i| + |y_i|) \stackrel{(i)}{\leq} \max_i (|x_i| + \max_j |y_j|) \\ &= \max_i |x_i| + \max_j |y_j| = \|x\|_\infty + \|y\|_\infty \end{aligned}$$

Therefore,  $\|x+y\|_\infty \leq \|x\|_\infty + \|y\|_\infty$

(3) definiteness

Assume  $\|x\|_\infty = \max_i |x_i| = 0 \Leftrightarrow \forall i |x_i| = 0 \stackrel{(iii)}{\Leftrightarrow} \forall i x_i = 0$   
 $\Leftrightarrow x = \vec{0}$

Therefore  $\|x\|_\infty = 0 \text{ iff } x = \vec{0}$

(4) nonnegativity.

$$\begin{aligned} 0 &= \|0\|_\infty = \|x + (-x)\|_\infty \leq \|x\|_\infty + \|-x\|_\infty = \|x\|_\infty + (-1) \|x\|_\infty \\ &= \|x\|_\infty + \|x\|_\infty \geq 2 \|x\|_\infty \end{aligned}$$

Therefore,  $0 \leq \|x\|_\infty$

3.24(a) let  $x = (1, 1)$ ,  $z_1 = (1, 0)$ ,  $z_2 = (5, 5)$

$$\angle(x, z_1) = \arccos\left(\frac{1}{\sqrt{2} \times 1}\right) = \arccos\left(\frac{\sqrt{2}}{2}\right) = \pi/4$$

$$\angle(x, z_2) = \arccos\left(\frac{10}{\sqrt{2} \times \sqrt{50}}\right) = \arccos\left(\frac{10}{\sqrt{2} \times \sqrt{2} \times 5}\right) = \arccos(1) = 0$$

$$\text{dist}(x, z_1) = \|x - z_1\| = \sqrt{(1-1)^2 + (1-0)^2} = \sqrt{1} = 1$$

$$\text{dist}(x, z_2) = \|x - z_2\| = \sqrt{(1-5)^2 + (1-5)^2} = \sqrt{32} = 4\sqrt{2}$$

Therefore,  $z_1$  is  $x$ 's closest distance neighbor, but  $z_2$  is  $x$ 's closest angle neighbor.

$$(b) \angle(x, z_i) = \arccos\left(\frac{x^T z_i}{\|x\| \|z_i\|}\right) = \arccos\left(\frac{x^T z_i}{\|x\|}\right)$$

$$= \arccos\left(\frac{\frac{1}{2}(\|x\| + \|z_i\|) - (\|x - z_i\|)}{\|x\|}\right)$$

$$= \arccos\left(\frac{\frac{1}{2}(\|x\| - \|x - z_i\| + 1)}{\|x\|}\right)$$

Since  $\arccos$  is a decreasing function as the numerator grows the angle decreases.

given  $u = (x, z_j) \nparallel v = (x, z_i) \forall i$

if  $\text{dist}(u) \leq \text{dist}(v)$

then  $\angle(u) \leq \angle(v)$ .

3.26(a)  $R(0) = 1$  because you are essentially finding the correlation between the same vector. A vector correlated with itself should be 1.

$R(T) = 0$  for  $T \geq T$  because the values past the time series are zero so you are finding the Correlation between time series  $x$  and the zero vector, which is equal to zero.

$$(b) R(T) = \rho(x_t, x_{t+T}) = \frac{1}{T} \sum_{t=1}^T z_t * z_{t+T} = \frac{1}{T} \left( \sum_{t=1}^{T-T} z_t * z_{t+T} + \sum_{t=T-T+1}^T z_t * z_{t+T} \right)$$

$$= \frac{1}{T} \left( \sum_{t=1}^{T-T} z_t * z_{t+T} + 0 \right) = \frac{1}{T} \sum_{t=1}^{T-T} z_t * z_{t+T}$$

Therefore,  $R(T) = \frac{1}{T} \sum_{t=1}^{T-T} z_t * z_{t+T}$

Since the correlation  $R(T) = 0$  for  $T \geq T$ .

$$(c) R(T) = (1, -0.99, 0.98, -0.97, \dots, 0.02, -0.01, 0)$$

(d) Since the meals served on the same day should be similar we should expect a high correlation between  $R(7) \& R(14)$  since they are on the same day, but we could expect some deviation if there are external forces like weather or traffic that might reduce the number of meals sold and in return reduce correlation.

### Problem A4.2 (Written Portions are in the comments of the Julia Screen Shots)

```
In [14]: # import files
include("wikipedia_corpus.jl")
include("kmeans.jl")

WARNING: replacing module Kmeans

Out[14]: Kmeans
```

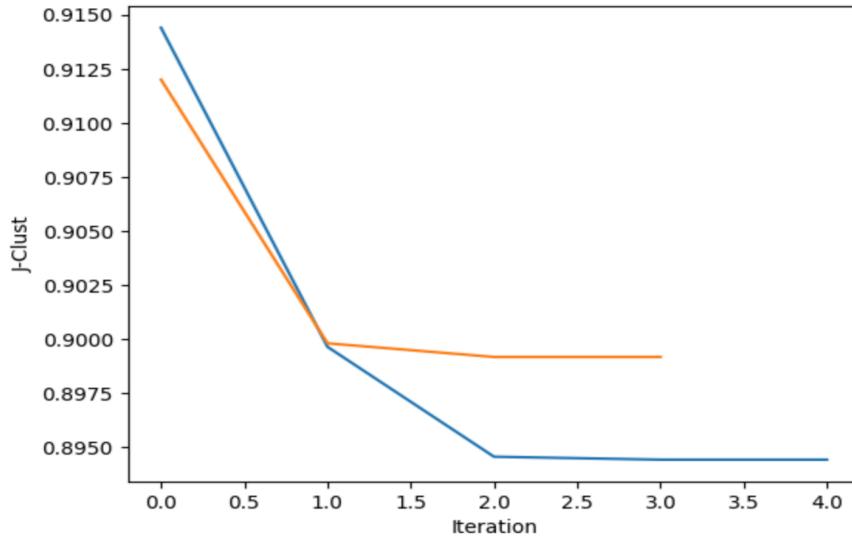
```
In [15]: using Kmeans
# 2-means
centroids1, labels1, j_hist1 = kmeans(article_histograms, 2)
centroids2, labels2, j_hist2 = kmeans(article_histograms, 2)

# 5-means
centroids3, labels3, j_hist3 = kmeans(article_histograms, 5)
centroids4, labels4, j_hist4 = kmeans(article_histograms, 5)

# 10-means
centroids5, labels5, j_hist5 = kmeans(article_histograms, 10)
centroids6, labels6, j_hist6 = kmeans(article_histograms, 10)
```

```
In [16]: # plot 2-means j_hist
using PyPlot
plt[:plot](j_hist1)
plt[:plot](j_hist2)
plt[:ylabel]("J-Clust")
plt[:xlabel]("Iteration")
plt[:show]

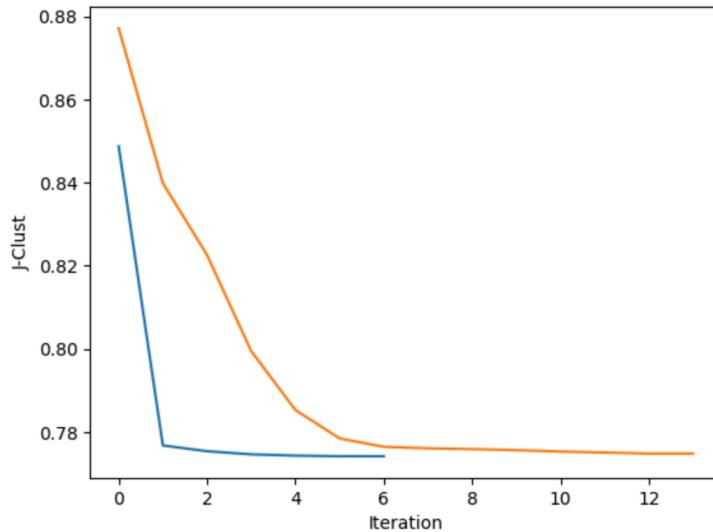
# The first run seems to have converged before the second run
# but forms a worse clustering since it has a larger Jclust.
```



```
Out[16]: PyObject <function show at 0x7f4640588668>
```

```
In [17]: # plot 5-means j_hist
plt[:plot](j_hist3)
plt[:plot](j_hist4)
plt[:ylabel]("J-Clust")
plt[:xlabel]("Iteration")
plt[:show]

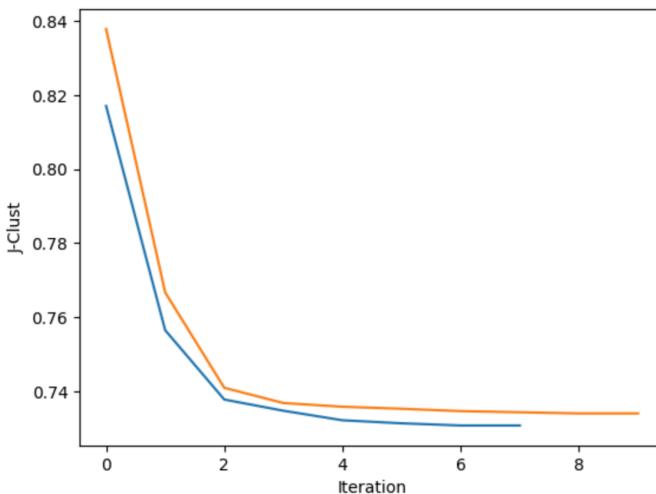
# The first run seems to converge in more iterations than run two
# but both runs have similar Jclusts which means that both runs
# have a similarly distributed end cluster. That doesn't necessarily
# mean both clusters are the same just that they have the same
# distribution of points.
```



```
Out[17]: PyObject <function show at 0x7f4640588668>
```

```
In [18]: # plot 10-means j_hist
plt[:plot](j_hist5)
plt[:plot](j_hist6)
plt[:ylabel]("J-Clust")
plt[:xlabel]("Iteration")
plt[:show]

# boths runs follow a very similar convergence path except for the
# fact that run one takes two more iterations. Both runs have very
# similar final Jclusts which means both have a similar end
# distribution.
```



```
In [34]: using Kmeans
centroids, labels, j_hist = kmeans(article_histograms, 5)

Out[34]: (Array{T,1} where T: [0.0113911, 0.00860466, 0.00840295, 0.00415809, 0.0028484, 0.0035467, 0.0117424, 0.00179013, 0.00345708, 0.00558935 ... 0.00239791, 0.00105711, 0.00278769, 0.00169994, 0.0, 0.010124, 0.00302327, 0.00151568, 0.000770211, 0.00532334], [0.00204671, 0.00600554, 0.00522378, 0.00504795, 0.00103929, 0.00335796, 0.00290721, 0.00285926, 0.00279841, 0.00328599 ... 0.00182123, 0.0025873, 0.00606146, 0.00159825, 0.0, 0.00263852, 0.000660748, 0.00225136, 0.00133602, 0.00447413], [0.0053282, 0.00737482, 0.00577754, 0.00390763, 0.00280889, 0.00208566, 0.0019377, 0.0023487, 0.0, 0.00210245 ... 0.000436471, 0.00171766, 0.00561198, 0.00708966, 0.0, 0.00488392, 0.0269847, 0.001829, 0.00446309, 0.00212861], [0.0102801, 0.00762798, 0.00287629, 0.00544805, 0.0056058, 0.00918726, 0.00725137, 0.00413262, 0.000498087, 0.00585914 ... 0.0127757, 0.0151872, 0.0069792, 0.00101831, 0.0241923, 0.0261655, 0.000247707, 0.011922, 0.00556012, 0.000442087], [0.00879426, 0.00466363, 0.00423245, 0.00289408, 0.00236804, 0.00368026, 0.00482446, 0.00728875, 0.0321111, 0.00199351 ... 0.016368, 0.0637438, 0.000179816, 0.0195024, 0.0, 0.0131261, 0.00314463, 0.00455395, 0.0204319, 0.000123461]], [5, 1, 1, 5, 2, 2, 2, 1, 2, 1 ... 2, 5, 4, 4, 4, 4, 4, 3], [0.89015, 0.827869, 0.809544, 0.789593, 0.779484, 0.775716, 0.774415, 0.774026])
```

```
In [35]: article_titles[labels == 1]

Out[35]: 99-element Array{String,1}:
"Acid rain"
"Albedo"
"Anemometer"
"Anticyclone"
"Atmosphere"
"Atmosphere of Earth"
"Atmospheric pressure"
"Baroclinity"
"Barograph"
"Barometer"
"Black ice"
"Ceiling balloon"
"Ceiling projector"
:
"United Nations Relief and Works Agency for Palestine Refugees in the Near East"
"Weather balloon"
"Weather forecasting"
"Weather front"
"Weather map"
"Weather modification"
"Weather radar"
"Wind chill"
"Wind direction"
"Windsock"
"Wind speed"
"Winter storm"
```

```
In [36]: dictionary[sortperm(centroids[1],rev=true)]
# the topic discovered seems to have something relating to weather and meterology.
# the top three words are weather, wind and pressure which in the context of meterology
# makes sense since all three words are heavily related to the atmosphere.
```

```
Out[36]: 1000-element Array{String,1}:
"weather"
"wind"
"pressure"
"air"
"temperature"
"ice"
```

```
In [37]: article_titles[labels == 2]
Out[37]: 58-element Array{String,1}:
"Amplitude modulation"
"Amplitude-shift keying"
"Analog signal"
"Antenna (radio)"
"Asynchronous Transfer Mode"
"Attenuation"
"Audio power amplifier"
"Broadcasting"
"Carrier signal"
"Channel (communications)"
"Communications satellite"
"Data transmission"
"Digital signal"
⋮
"Smoke signal"
"Telecommunications network"
"Telephone"
"Telephone exchange"
"Telepresence"
"Time-division multiplexing"
"Transmission medium"
"Transmitter"
"Voice over IP"
"W-CDMA (UMTS)"
"Wide area network"
"Wireless"
```

```
In [38]: dictionary[sortperm(centroids[2],rev=true)]
# the topic discovered seems to have something relating to communication and sound.
# the top three words are signal, radio and frequency which in the context of sound and communication
# makes sense since radios work off of signals and frequencies to allow people to communicate.
```

```
Out[38]: 1000-element Array{String,1}:
"signal"
"radio"
"frequency"
"telephone"
"digital"
"transmission"
```

```
In [39]: article_titles[labels == 3]
Out[39]: 32-element Array{String,1}:
"Brock (Pokémon)"
"Bulbasaur"
"Deoxys"
"Eevee"
"Gameplay of Pokémon"
"Hey You, Pikachu!"
"Lapras"
"List of Pokémon characters"
"Meowth"
"Mew (Pokémon)"
"Mewtwo"
"Misty (Pokémon)"
"Pikachu"
⋮
"Pokémon HeartGold and SoulSilver"
"Pokémon Platinum"
"Pokémon Red and Blue"
"Pokémon Ruby and Sapphire"
"Pokémon Trading Card Game"
"Pokémon universe"
"Pokémon Yellow"
"Pokémon 4Ever"
"Satoshi Tajiri"
"Team Rocket (anime)"
"Togepi"
"Zapdos"
```

```
In [40]: dictionary[sortperm(centroids[3],rev=true)]
# the topic discovered seems to have something relating to pokémon.
# the top three words are pokémon, game and games which in the context
# of pokémon makes sense since pokémon is a video game and anime. The problem
# with this cluster seems to be that the lack of clusters have made articles
# like wavelengths and wind which could most likely belong to clusters two
# and one respectively. The overall most frequent words are definitely relevant.
```

```
Out[40]: 1000-element Array{String,1}:
"pokémon"
"game"
"games"
```

```
In [41]: article_titles[labels == 4]

Out[41]: 65-element Array{String,1}:
  "Convention on the Rights of Persons with Disabilities"
  "Food and Agriculture Organization"
  "Headquarters of the United Nations"
  "International Bank for Reconstruction and Development"
  "International Centre for Settlement of Investment Disputes"
  "International Civil Aviation Organization"
  "International Court of Justice"
  "International Development Association"
  "International Finance Corporation"
  "International Fund for Agricultural Development"
  "International Labour Organization"
  "International Maritime Organization"
  "International Monetary Fund"
  :
  "United Nations Trusteeship Council"
  "United Nations University"
  "United Nations Volunteers"
  "Universal Postal Union"
  "UN Women"
  "World Bank Group"
  "World Food Programme"
  "World Health Organization"
  "World Intellectual Property Organization"
  "World Meteorological Organization"
  "World Tourism Organization"
  "World Trade Organization"
```

```
In [42]: dictionary[sortperm(centroids[4],rev=true)]
# the topic discovered seems to have something relating to international affairs.
# the top three words are nations, international and member. All of these words are
# very connected to the topic of international affairs. One of the words in actually in
# the name of the topic. As we go down the list we start to see that the other more popular
# make sense give the context.
```

```
Out[42]: 1000-element Array{String,1}:
  "nations"
  "international"
  "member"
  "council"
```

```
In [43]: article_titles[labels == 5]

Out[43]: 46-element Array{String,1}:
  "A Bar at the Folies-Bergère"
  "Alfred Sisley"
  "Armand Guillaumin"
  "Art Institute of Chicago"
  "A Sunday Afternoon on the Island of La Grande Jatte"
  "Bal du moulin de la Galette"
  "Berthe Morisot"
  "Café Guerbois"
  "Camille Pissarro"
  "Claude Monet"
  "Cubism"
  "Édouard Manet"
  "Dutch Golden Age painting"
  :
  "Neo-impressionism"
  "Paris Street; Rainy Day"
  "Paul Cézanne"
  "Paul Signac"
  "Pierre-Auguste Renoir"
  "Portrait"
  "Post-Impressionism"
  "Salon des Refusés"
  "Still life"
  "The Child's Bath"
  "Wet-on-wet"
  "Woman with a Parasol - Madame Monet and Her Son"
```

```
In [44]: dictionary[sortperm(centroids[5],rev=true)]
# the topic discovered seems to have something relating to art.
# the three most popular words are painting, art and paintings which
# make sense in the context of art. Additionally, looking further into
# the list we see a lot of the other words make sense in the topic
# of art.
```

```
Out[44]: 1000-element Array{String,1}:
  "painting"
  "art"
  "paintings"
  "artists"
```