Week 8　Quiz 1　　1回目　/

X 1. k-means クラスタリングが適切なアルゴリズムである場合を選べ

① ユーザ情報のデータベースが与えられた時，異なる marketing segment に自動的にグループ分けしろ　○

② スーパーマーケットのたくさんの品物の販売記録から，同時にかわれているものをグループ分けして同じ棚に並べられたい．　X といてみる　X　○が正解であった．

k-means には任意の「2点間の距離」が定義できる必要がある．
同時に売れた回数の逆数とかを距離にできるか？　加法定理がなりたたないので距離ではない

③ 過去の天気の記録から，明日の雨量を数値で predict する．

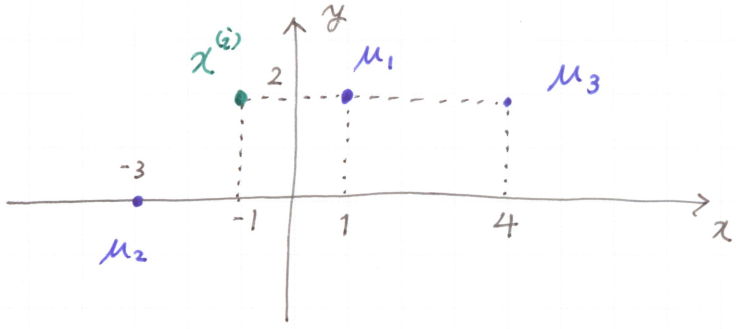
× k グループに分ける必要がない．
今日の天気がどこにグループ化されるかわかっても 雨量の数値はわからない
クラスタの平均？
最小？
最大？
最頻？

④ スーパーマーケット　：　から，各商品の今後の売れ方を予測する　X

Week 8　Quiz 1　1回目　2

2

$\mu_1 = \begin{pmatrix}1\\2\end{pmatrix}$, $\mu_2 = \begin{pmatrix}-3\\0\end{pmatrix}$, $\mu_3 = \begin{pmatrix}4\\2\end{pmatrix}$, $x^{(i)} = \begin{pmatrix}-1\\2\end{pmatrix}$



最も $x^{(i)}$ に近いのは μ_1　∴ $c^{(i)} = \mu_1$ の index $= 1$

3

K-means は 2つの step が くり返し実行される。2つの step を選べ。

① クラスタ割り当て step で $c^{(i)}$ が更新される　○

② cluster centroid μ_j を動かす　○

③ μ_K を動かす（最も近い $x^{(i)}$ に重なるように）　×

④ クラスタ割り当て step で μ_j を最も近い $x^{(i)}$ に重なるように μ_j を動かす　×

4

unlabeled dataset $\{x^{(1)}, \cdots, x^{(m)}\}$, 50回の random initialization で 50種のクラスタリング → どの分類結果を選ぶべきか？

① ない　×

② ラベルがある時だけできるので、無理　×

③ $\frac{1}{m}\sum_{i=1}^{m} \| x^{(i)} - \mu_{c^{(i)}} \|^2$ が最小のものを選ぶ　○

④ 50回目の結果を選べ。収束していくので　×（最後の）

Week 8　Quiz 1　1回目　3

5

① 初期値として $\mu_1 = \cdots = \mu_K = 0$ を選べ　×

② 「正しい」または「適切な」Kの値があいまいなデータセットもある。
人間のエキスパートでもなかなか決定できない。　○

③ local optimum に落ちることを心配するまでは，複数回 random initialization を使え　○

④ K-means は unsupervised なので，data に overfit することはない。
計算可能なかぎりクラスター数は多い方がいい。　×

Week 8　Quiz 1　　2回目

1

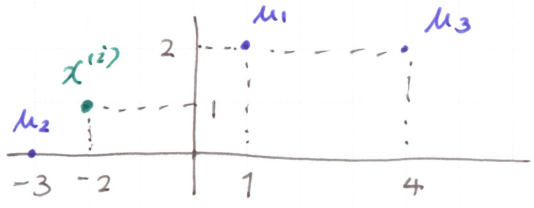
① ユーザ情報から　○

② スーパーで同時に売れたもの　○

③ 過去の天気から明日の雨量　×

④ スーパー　今後の売れ行き　×

2



$c^{(i)} = 2$ ($= \mu_2$ の index)

3　k-means に含まれる2つの step

① K を選ぶための elbow method ×

② クラスタ割り当て step で $c^{(i)}$ を更新　○

③ cluster centroid 移動 step, μ_k を動かす　○

④ feature scaling　×

4　unlabeled $\{x^{(1)}, \cdots, x^{(m)}\}$, 50回クラスタ4 → 50個の結果

① $\frac{1}{m}\sum_i \| x^{(i)} - \mu_{c^{(i)}} \|^2$ 最小をえらぶ

5

① $\mu_1 = \cdots = \mu_K = 0$ を初期値　×

② local optima を避けるため初期化を複数回　○

③ unsupervised なので overfit はよい、クラスタ数はできるだけ多く　×

④ どっちが正しいかわからないデータもある.