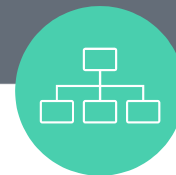
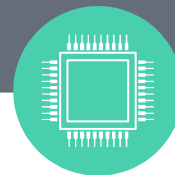


مقدمه‌ای بر یادگیری ماشین و یادگیری ژرف

محمد حسین رهبان
جلسه اول

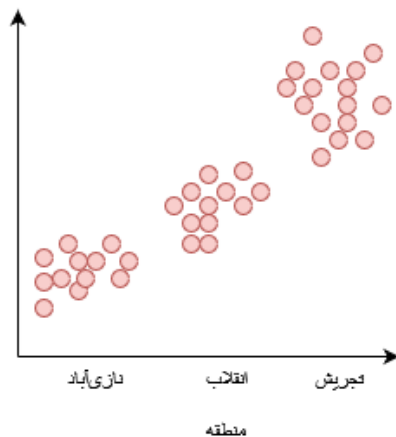
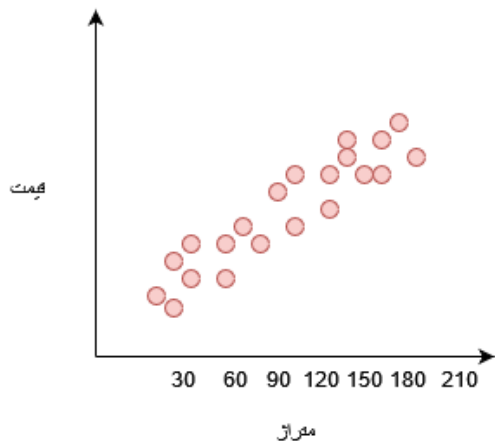


اطلاعات درس

- مخاطبین: برنامه نویسی مقدماتی (پایتون)
- ساختار: ۸ جلسه – Hands on و عملیاتی
- وبسایت: اطلاع رسانی می شود (کوئرا)



پیش‌بینی قیمت ملک



فاکتورهای مهم؟

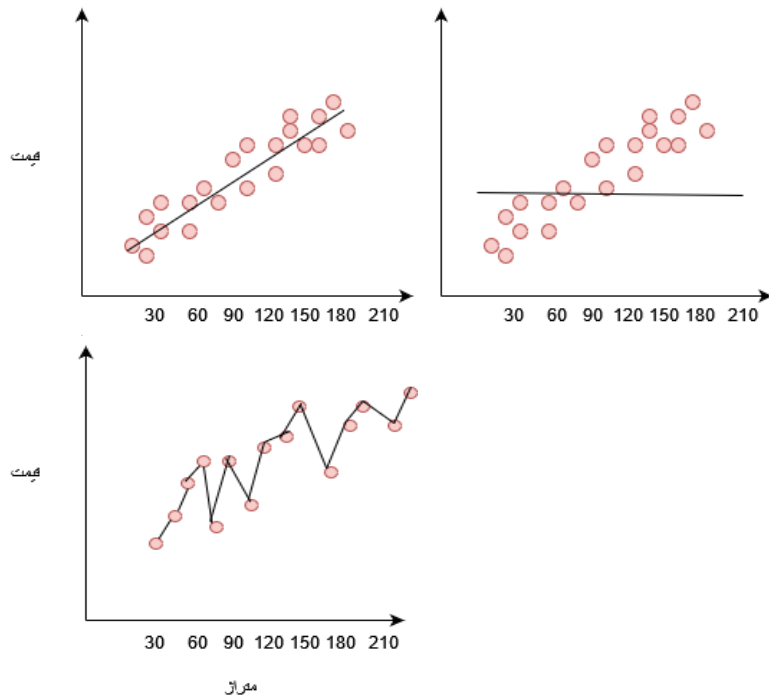
- منطقه یا محله
- متراژ
- پارکینگ
- تعداد واحد
- تعداد طبقه



AI CONTEST

rayan

پیش‌بینی قیمت ملک



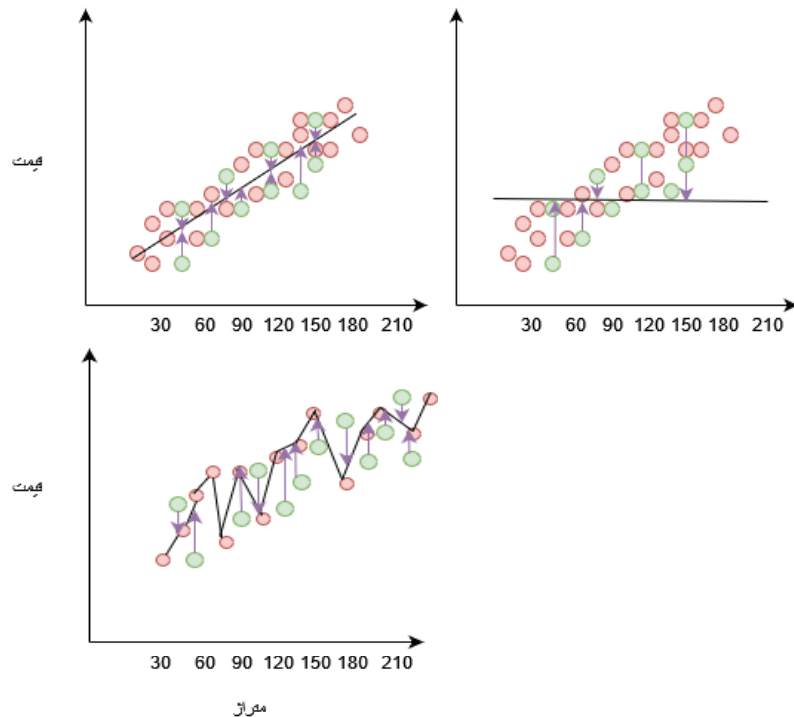
چگونه قیمت ملک‌هایی که **ندیده‌ایم** را تخمین بزنیم؟



AI CONTEST

rayan

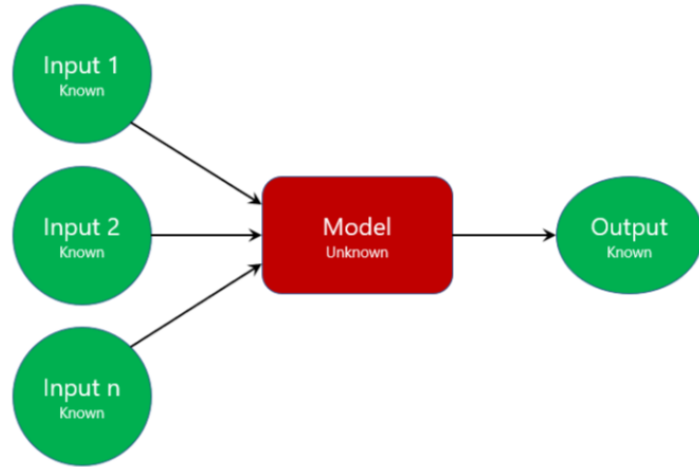
پیش‌بینی قیمت ملک



کدام مدل‌سازی مناسب است؟

- کدام مدل‌سازی اصلاً داده‌ها را یاد نگرفته؟ (Underfit)
- کدام مدل روی داده‌ی جدید عملکرد خوبی ندارد؟ (Overfit)

یادگیری ماشین (machine learning)



- تخمین یک متغیر (قیمت ملک)
- داده
- تعمیم (Generalization)
- نظم، الگو، یا قاعده

یک چالش بامزه

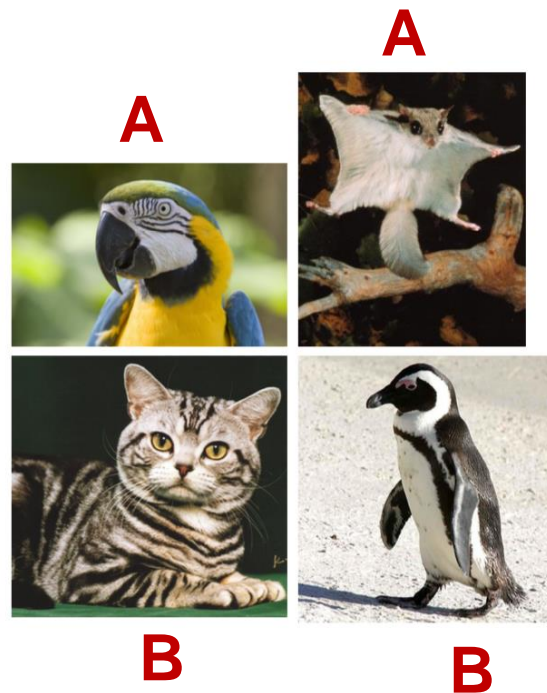
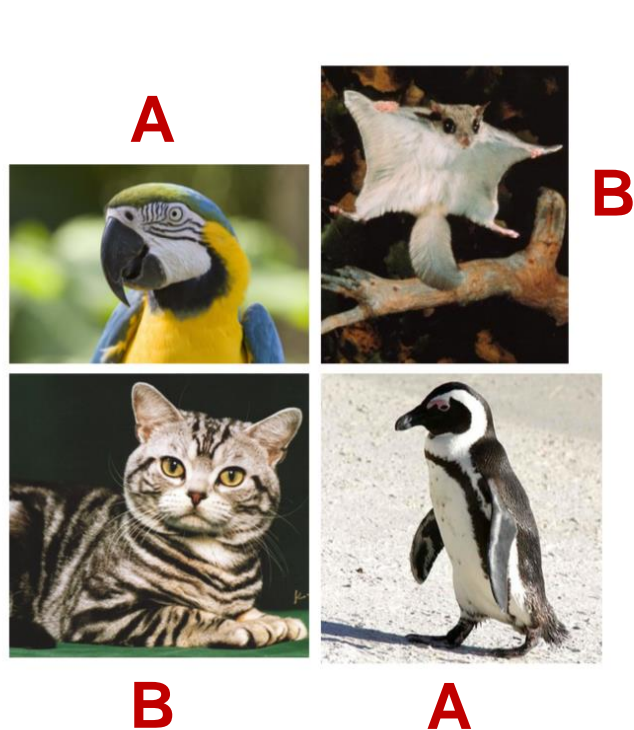
داده‌ها

class A

class B

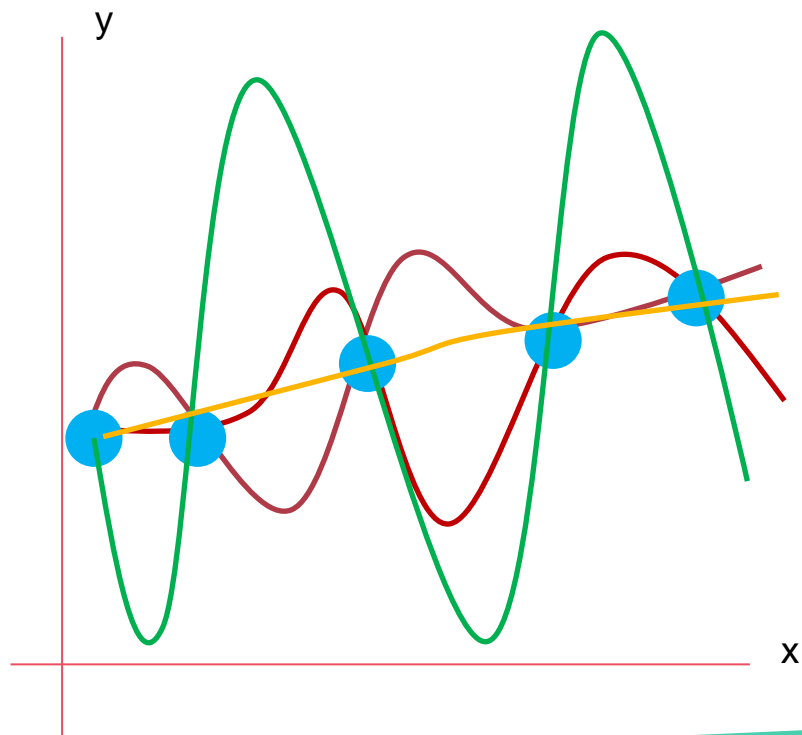


کدام برچسب‌زنی صحیح است؟



Courtesy: Hal Daumé III, A Course in Machine Learning, 2017

کدام منحنی صحیح است؟



AI CONTEST

rayan

انواع یادگیری ماشین

یادگیری نظارتی (Supervised Learning)

- پیش‌بینی خروجی بر اساس داده‌های برچسب‌گذاری شده (رگرسیون و دسته‌بندی)

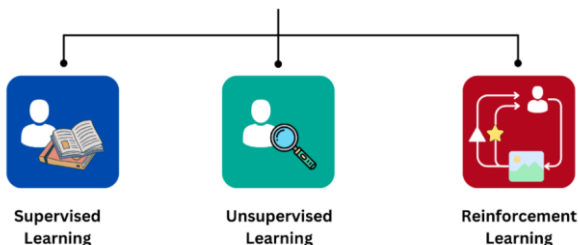
یادگیری بدون نظارت (Unsupervised Learning)

- پیدا کردن الگوها و ویژگی‌های مخفی در داده (خوشه‌بندی و کاهش بعد)

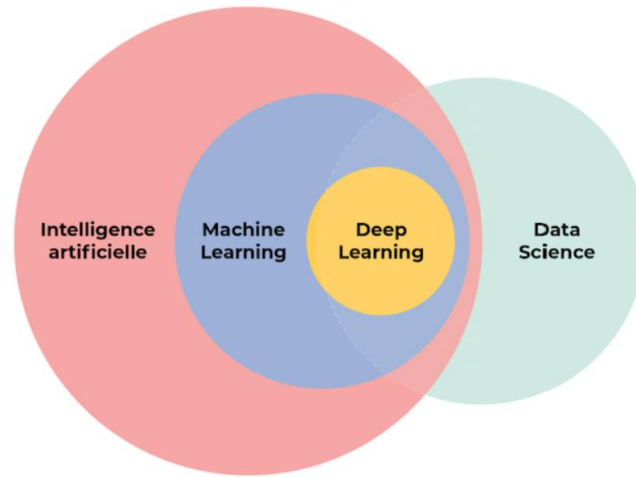
یادگیری تقویتی (Reinforcement Learning)

- یادگیری عمل یا فعل بر اساس پاداش در طول زمان (رباتیکس)

Types of Machine Learning

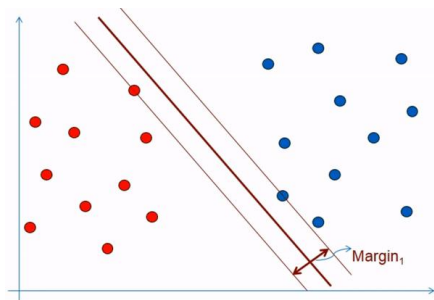


هوش مصنوعی، یادگیری ماشین، یادگیری عمیق، علوم داده

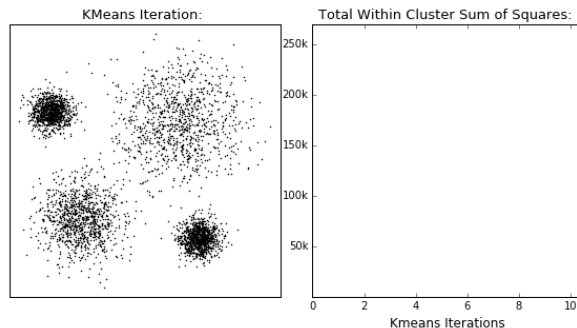


مثال‌هایی از یادگیری ماشین

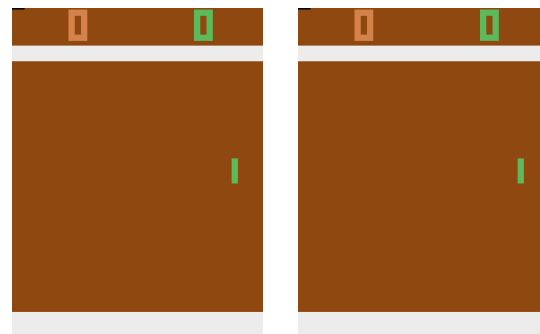
یادگیری نظارتی



یادگیری بدون نظارت



یادگیری تقویتی

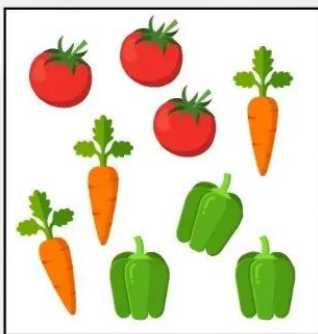


AI CONTEST

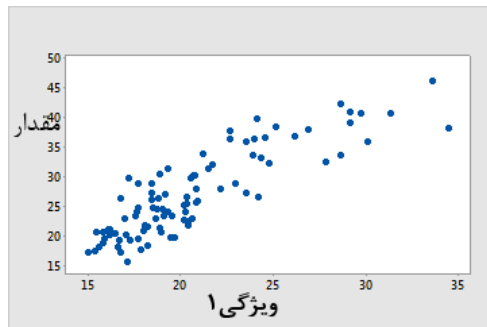
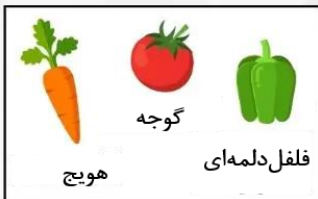
rayan

الگوریتم‌های کلیدی

داده برچسب گذاری شده



برچسب‌ها



یادگیری نظارتی

• Logistic Regression و Linear Regression

• Decision Tree

• Support Vector Machines (SVM)

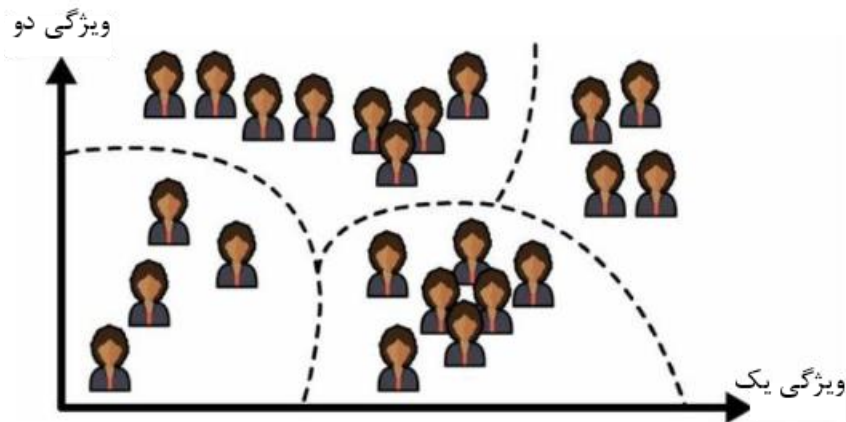
• Neural Networks



AI CONTEST

rayan

الگوریتم‌های کلیدی



یادگیری بدون نظارت

- K-means Clustering
- Hierarchical Clustering
- PCA
- t-SNE



AI CONTEST

rayan

رگرسیون خطی

- مدل کردن رابطه بین خروجی و ورودی
- پیدا کردن بهترین (کم هزینه‌ترین) خط یا صفحه برای دادگان

معادله:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

هدف: یافتن بهترین $\beta_{1..n}$



AI CONTEST

rayan

رگرسیون خطی

• تابع هزینه

- MSE: $\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$
- MAE: $\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$

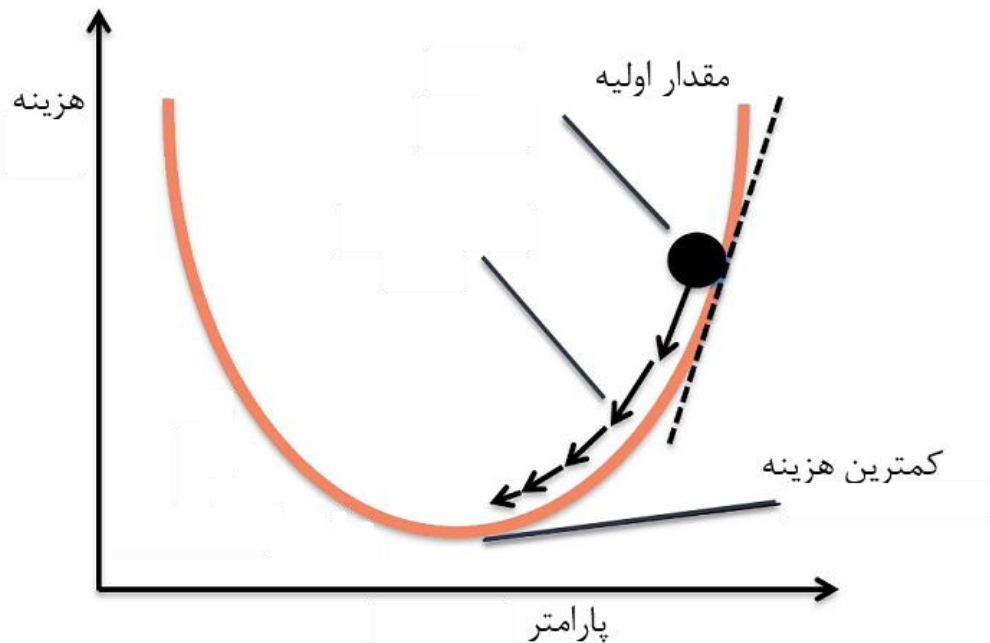


AI CONTEST

rayan

رگرسیون خطی

گرادیان کاهشی



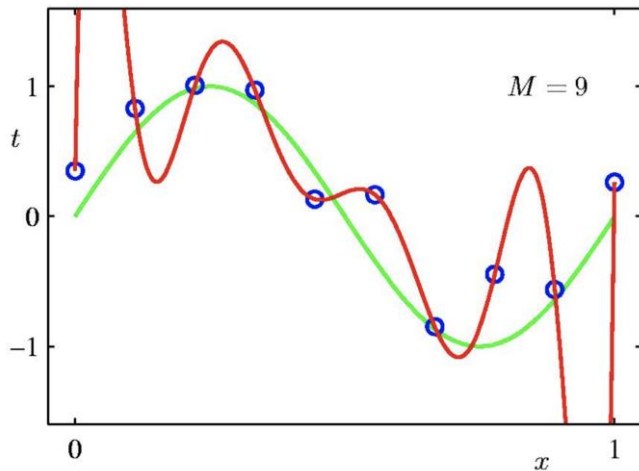
مثال تعاملی: رگرسیون خطی



AI CONTEST

rayan

بیش برزش (overfitting)



- مدل پیچیده: چند جمله‌ای درجه ۹ به جای خط
- خطا روی داده آموزشی پایین
- خطا روی داده دیده نشده (آزمایشی) بالا
- چگونه رفع کنیم؟ داده بیشتر – مدل ساده‌تر – منظم‌سازی (regularization)



K-Means

مفهوم:

- روشی برای دسته بندی و تقسیم بندی داده ها به خوشه های مجزا.

الگوریتم:

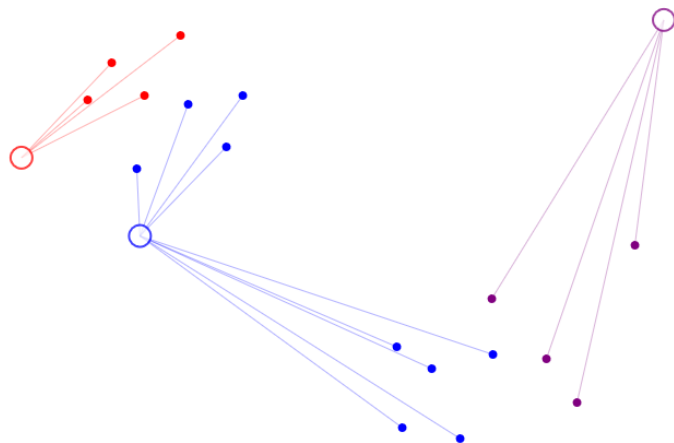
- قرار دادن تعدادی مرکز خوشه بصورت تصادفی.
- تخصیص هر داده به یک مرکز بر اساس فاصله اقلیدسی.
- بروزرسانی مرکز خوشه ها بر اساس میانگین داده های داخل هر خوشه.
- تکرار مراحل بالا تا همگرایی.



AI CONTEST

rayan

K-Means



- قرار دادن تعدادی مرکز خوشه بصورت رندم.
- تخصیص هر داده به یک مرکز بر اساس فاصله اقلیدسی.

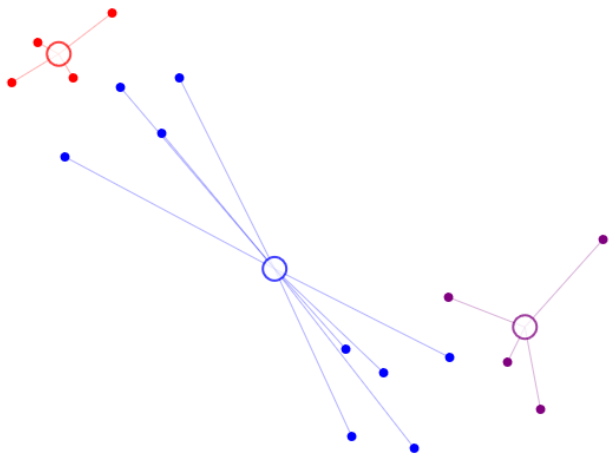
چگونه تعداد مراکز خوشه را مشخص کنیم؟
آیا محل اولیه مراکز خوشه تاثیری در همگرایی دارند؟



AI CONTEST

rayan

K-Means



- بروزسانی مرکز خوشه‌ها بر اساس میانگین داده‌های داخل هر خوشه.



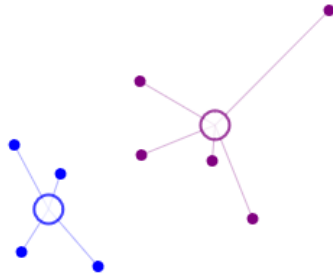
AI CONTEST

rayan

K-Means



● تکرار مراحل



مثال تعاملی KNN



AI CONTEST

rayan