ВЕКТОРЫ В МАШИННОМ ОБУЧЕНИИ

INNOSTAGE

О ЧЕМ ПОЙДЕТ РЕЧЬ

01	Введение в векторы
02	Представление векторов - базис, координаты, разложение по базису
03	Метрики для векторов
04	Векторы для анализа данных - PCA
05	Векторы в машинном обучении

Векторы

INNOSTAGE

01 Аналитическая геометрия и векторы

01.1 - Геометрическое представление

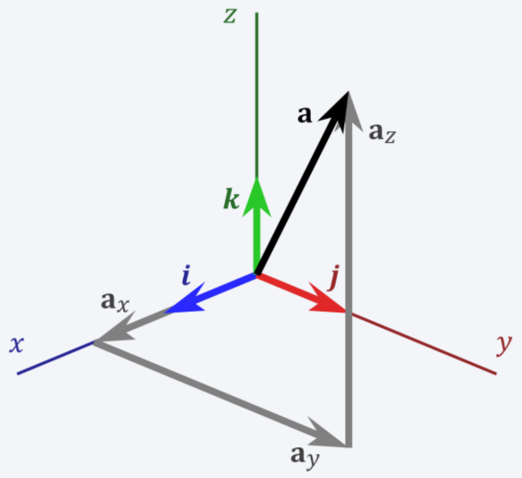
01.2 - Линейные операции

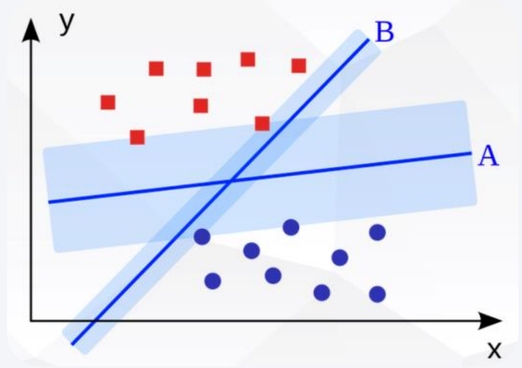
02 Векторы в машинном обучении

02.1 - Представление данных

02.2 - Параметры моделей

02.3 - Вычисления и прогнозы





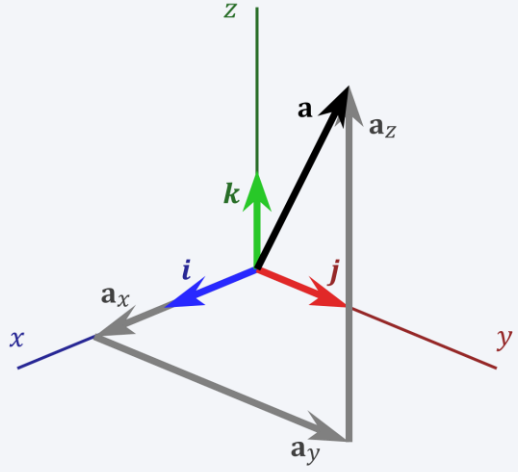
Свойства векторов

INNOSTAGE

01

Модуль (Длина)

Модуль вектора представляет собой его длину и обозначается обычно как $|v|$.



02

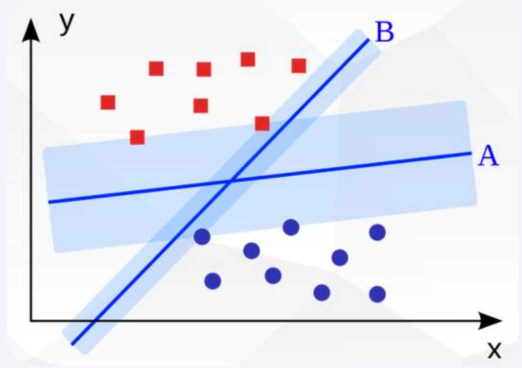
Направление

Направление вектора определяется углом, который он образует с определенной осью или другим вектором.

03

Координаты в Выбранной Системе

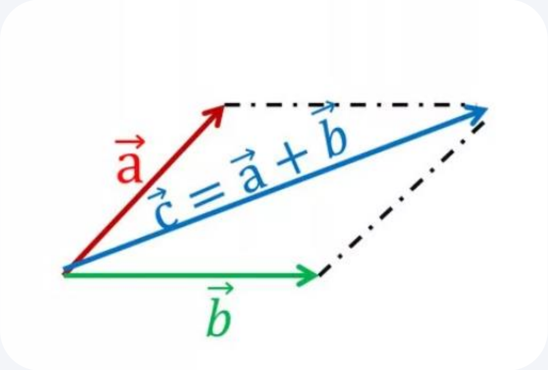
Для более удобного анализа и работы с векторами их можно представить в виде набора координат в выбранной системе координат



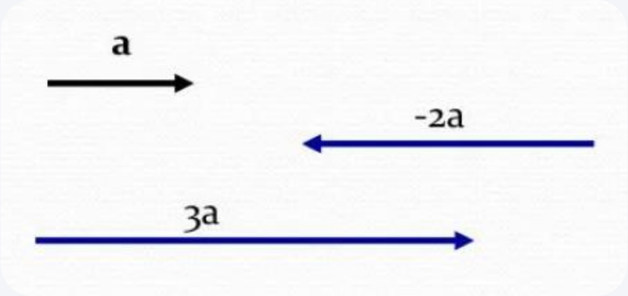
Операции над векторами

INNOSTAGE

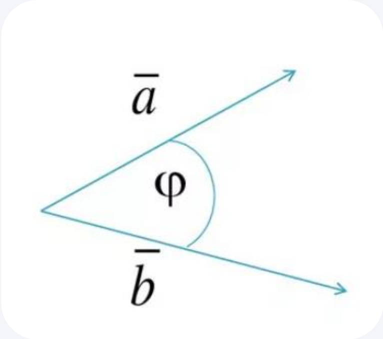
01 Сложение векторов (по правилу параллелограмма)



02 Умножение вектора на число



03 Скалярное произведение

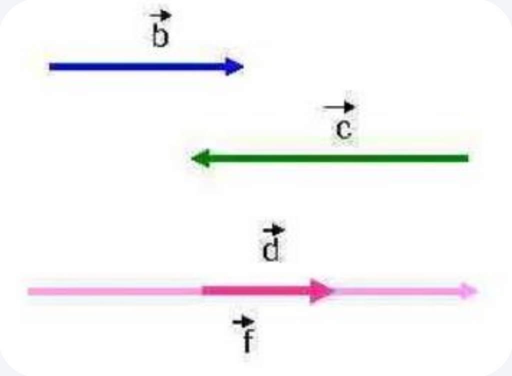


Коллинеарность и компланарность векторов

INNOSTAGE

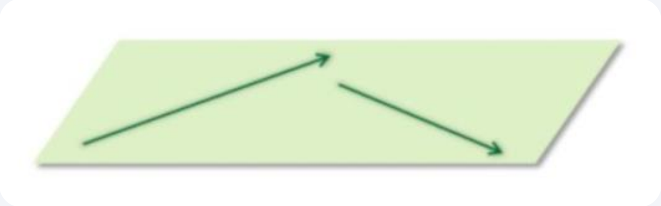
01 Коллинеарные векторы

Векторы, которые лежат на одной и той же прямой



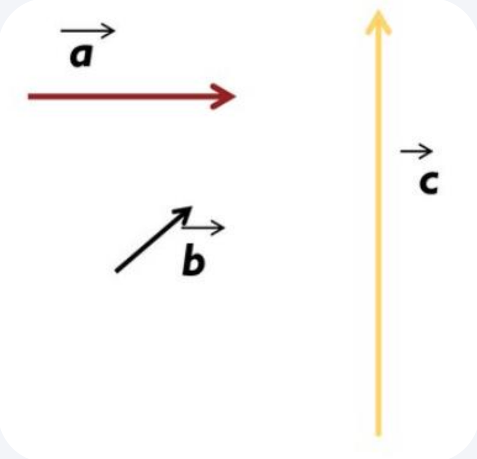
02 Компланарные векторы

Векторы, которые лежат в одной и той же плоскости



03 Некомпланарные векторы

Векторы, которые не лежат в одной и той же плоскости



Коллинеарность и компланарность векторов

INNOSTAGE

01 Коллинеарные векторы

01.1 - Линейная зависимость

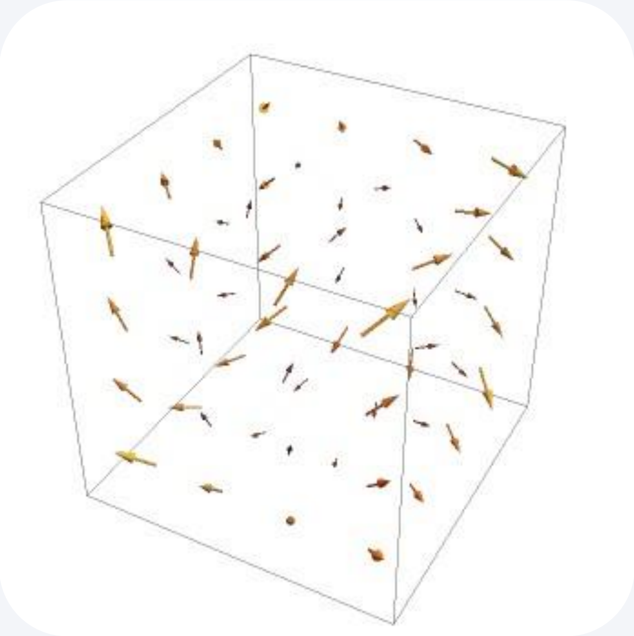
01.2 - Уменьшение размерности

02 Компланарные векторы

02.1 - Линейная зависимость в плоскости

02.2 - Проекция

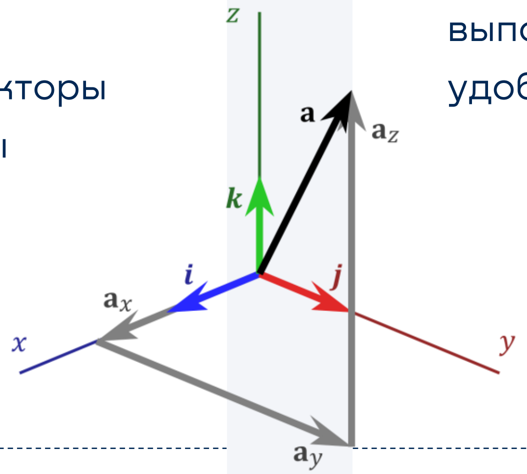
02.3 - Работа с плоскостями



Базис и координаты вектора

Базис определяет способ разложения любого вектора в данном пространстве на его составляющие векторы базиса.

Координаты вектора однозначно определяют положение этого вектора в данном базисе. Когда вектор разлагается по базису, его координаты представляют собой скалярные значения, которые умножаются на соответствующие векторы базиса и затем складываются, чтобы получить исходный вектор.



Координаты вектора представляют собой важную информацию о его расположении и направлении в данном пространстве. Они позволяют нам проводить анализ и выполнение операций над векторами в удобной координатной системе.

Разложение вектора по базису

Для того чтобы более глубоко понять структуру и относительное положение векторов в пространстве, мы можем представить любой вектор **a** в виде линейной комбинации базисных векторов:

$$a = a_1\boldsymbol{e_1} + a_2\boldsymbol{e_2} + \cdots + a_n\boldsymbol{e_n}$$

Где:

- **a** - исходный вектор, который мы хотим разложить.

- a1, a2, ..., an - коэффициенты, которые определяют, какую часть каждого базисного вектора следует взять для построения вектора a. Эти коэффициенты называются координатами вектора a в данном базисе.

- **e1, e2, ..., en**- базисные векторы, которые образуют базис пространства.

Значение в Data Science:

- Понимание структуры данных
- Извлечение признаков
- Снижение размерности данных
- Кластеризация и классификация

Стандартный базис

INNOSTAGE

На этом слайде мы рассмотрим стандартный базис в n-мерном пространстве R^n и его важное значение при работе с векторами в этом пространстве.

$$[e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}]$$

Для n-мерного пространства R^n стандартный базис включает в себя n векторов, где каждый вектор представляет собой единичный вектор вдоль одной из координатных осей.

- Стандартный базис и координаты вектора в нем являются ключевыми концепциями при работе с данными и векторными представлениями. Они позволяют нам легко анализировать и представлять данные в многомерных пространствах, а также проводить операции линейной алгебры.

Метрики расстояний

- Евклидова метрика (Евклидово расстояние)

$$d(u,v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}$$

- Манхэттенская метрика (Манхэттенское расстояние)

$$d(u,v) = \sum_{i=1}^{n}|u_i - v_i|$$

- Косинусная метрика (Косинусное расстояние)

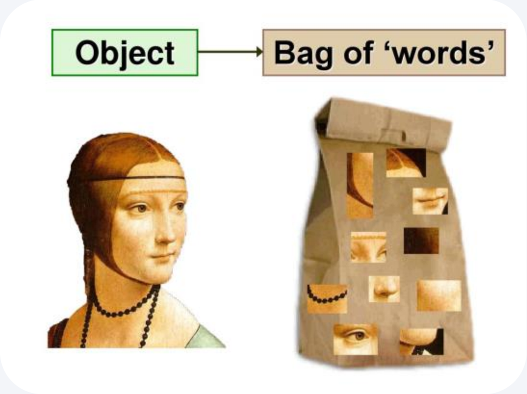
$$d(u,v) = \arccos\left(\frac{u \cdot v}{|u||v|}\right)$$

Примеры использования метрик

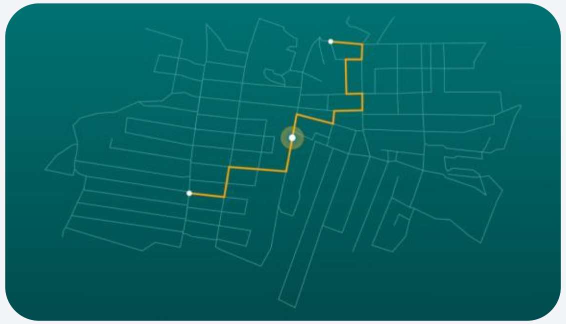
INNOSTAGE

01 **Евклидова** - Рекомендательные системы

02 **Косинусная** - Анализ текстовых данных



03 **Манхэттенская** - Планирование маршрутов в городской навигации



Метод главных компонент (PCA)

Метод главных компонент (PCA):

PCA - это метод, который помогает найти новый базис в пространстве признаков, такой, что проекции данных на этот базис максимально сохраняют дисперсию, а следовательно, информацию.

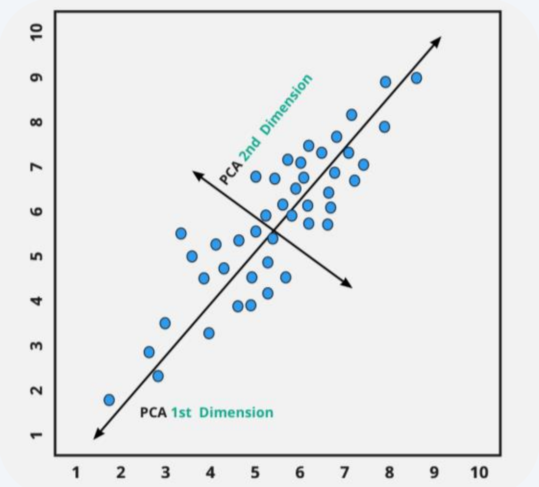
01 Центрирование данных

02 Вычисление ковариационной матрицы

03 Нахождение главных компонент

04 Проецирование данных

05 Выбор количества компонент



Для самостоятельного изучения:

- Математика PCA
- SVD и связь с PCA

Метод главных компонент (PCA)

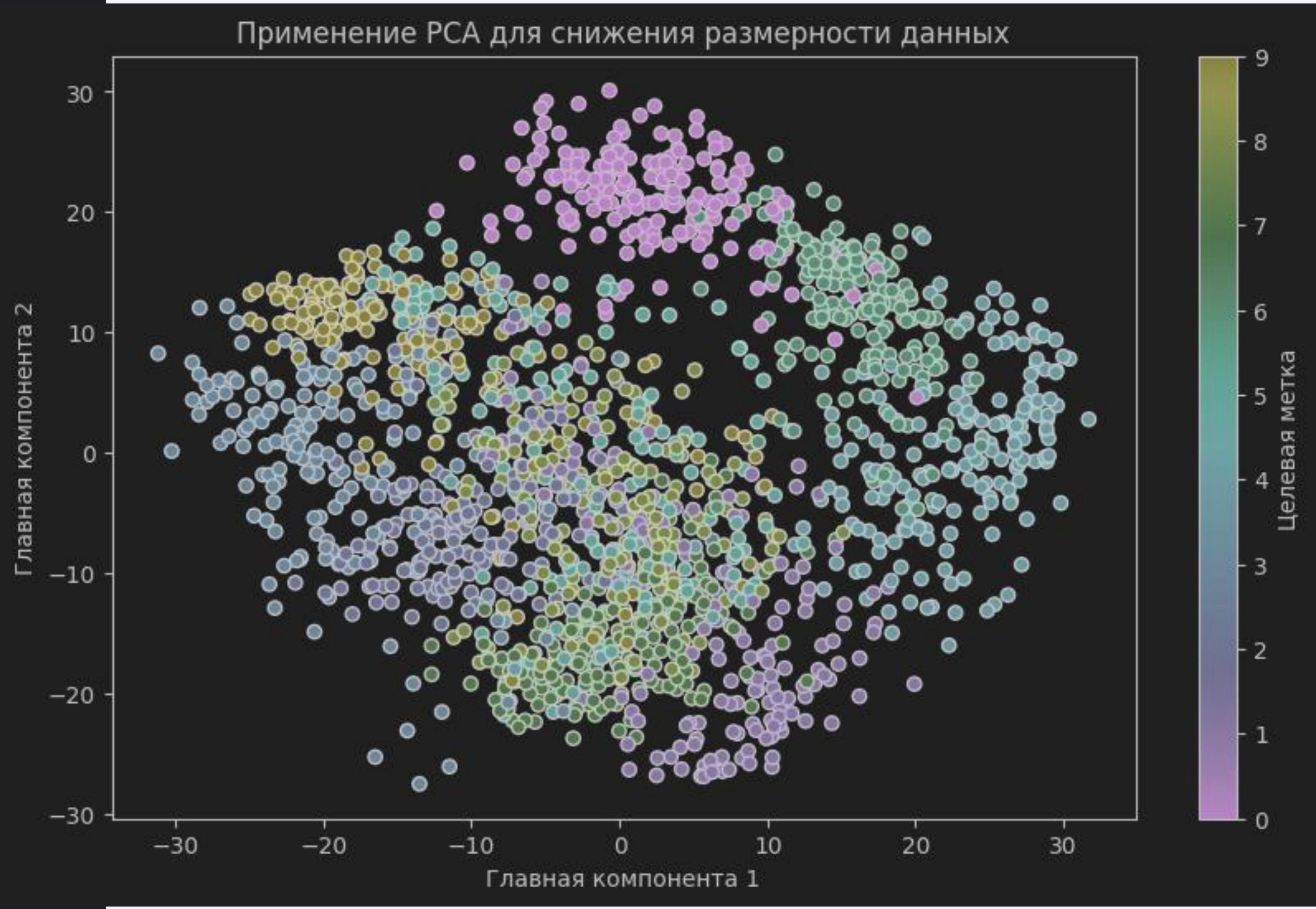
Пример: Анализ изображений

INNOSTAGE

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.datasets import load_digits

# Загрузка набора данных изображений рукописных цифр
digits = load_digits()
X = digits.data  # Векторы признаков
y = digits.target  # Метки классов
print(X.shape)
# Применение PCA для снижения размерности
pca = PCA(n_components=2)  # Указываем, что хотим снизить размерность до 2 компонент
X_pca = pca.fit_transform(X)

# Визуализация результатов
plt.figure(figsize=(10, 6))
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=y, cmap='viridis', edgecolor='k')
plt.colorbar(label='Целевая метка')
plt.xlabel('Главная компонента 1')
plt.ylabel('Главная компонента 2')
plt.title('Применение PCA для снижения размерности данных')
plt.show()
```



Выводы

INNOSTAGE

01 Векторы - мощный математический инструмент, широко используемый в различных областях, включая линейную алгебру, аналитическую геометрию и машинное обучение.

02 Векторы позволяют представлять данные и параметры моделей, что критически важно для эффективной работы с большими объемами данных.

03 Основные свойства векторов - это их длина, направление и координаты. Понимание этих свойств необходимо для глубокого анализа данных.

04 Операции над векторами, такие как сложение, умножение на число и скалярное произведение, являются фундаментальными инструментами линейной алгебры и machine learning.

05 Концепции базиса и координат позволяют разложить векторы для анализа их компонент. Это критически важно для понимания данных.

06 Правильный выбор метрики (евклидова, манхэттенская, косинусная) важен для решения конкретных задач анализа данных.

07 Методы вроде PCA используют разложение векторов по базису для снижения размерности и выделения важных признаков.

СПАСИБО ЗА ВНИМАНИЕ

INNOSTAGE

Казань, ул. Подлужная, 60

+7 (843) 567-42-90

info@innostage-group.ru