**1주차 Linear Algebra: for machine learning**

**학습 목표**

* Recall how machine learning and vectors and matrices are related
  + 기계학습: 특정 작업에 대한 데이터를 통해 기계가 특정 작업을 해낼 수 있게 훈련시키는 것
  + 데이터를 설명할 방정식을 계산할 때 필요
  + 벡터와 행렬은 데이터의 변수와 특성값(독립변수 - 딸기, 바나나 / 특성값 - 독립변수의 개수 / 종속변수 - 가격)을 표현
  + 즉, 데이터를 표현하고 연산할 때 사용하는 도구
* Interpret how changes in the model parameters affect the quality of the fit to the training data
  + 모델 파라미터에 따른 잔차 그래프를 확인
  + 실제로는 잔차 그래프를 알기 어렵기 때문에, **벡터의 기울기를 통해 더 낮은 잔차를 가지는 모델 파라미터 값을 찾는 것이 모델 파라미터 튜닝의 핵심**.
* Recognize that variations in the model parameters are vectors on the response surface - that vectors are a generic concept not limited to a physical real space
  + 벡터 공간은 데이터 혹은 모델 파라미터의 방향(변화)을 설명해줄 수 있음
* Use substitution / elimination to solve a fairly easy linear algebra problem
  + 연립 (선형) 방정식
* Understand how to add vectors and multiply by a scalar number
  + Multiple
  + 방향으로 접근: 벡터 공간에서 n배 만큼 반복해서 움직이면 됨. 단, 음수배의 경우 반대방향
  + 각 요소 별 접근: 벡터의 요소 별로 n배를 곱해줌.

**Introduction: Solving data science challenges with mathematics**

The purpose of this specialization is to take you on **a tour through the basic maths underlying these methods(machine learning)**, **focusing in particular on building your intuition** rather than worrying too much about the details.

it's actually possible to apply many powerful machine learning methods without understanding very much about the underpinning mathematics, by using open source libraries. This is great, **but problems can arise and without some sense of the language and meaning of the relevant maths, you can struggle to work out what's gone wrong or how to fix it.**

**This first course offers the introduction to linear algebra which is essentially a set of notational conventions and handy operations, that allow you to manipulate large systems of equations conveniently.**(선형 대수는 큰 방정식 체계를 편리하게 다룰 수 있도록 도와주는 표기 규칙과 유용한 연산의 집합입니다.)

we'll be focusing on building your intuition **about vectors and translations**…

**In the final module,** Dr. Sam Cooper will bring it all together by **showing you how linear algebra is at the heart of Google's famous page rank algorithm, which is used for deciding the order of web pages in search results.** Hopefully, if you find this course useful, you'll stick around for **a follow-on course by Sam and I who will introduce you to multivariate calculus.** Then, in our other course Dr. Mark Dyes and I will introduce principal component analysis.

**Motivations for linear algebra**

**-** the types of problems we might want to solve

1. The first problem I might think of is one of price discovery.

화이트보드, 친필, 텍스트, 폰트이(가) 표시된 사진

자동 생성된 설명

* ( ): 행렬, [ ]: 벡터

1. Another type of problem we might be interested in is fitting an equation to some data.

**Getting a handle on vectors**

**벡터가 쓰이는 문제:** 각 독립변수에 대해 선형인 경우에 가장 적합한 파라미터(매개변수)를 찾는 것

* 사람들의 키 분포

텍스트, 친필이(가) 표시된 사진

자동 생성된 설명

μ: 중앙값, σ: 얼마나 퍼져있는지

위는 정규분포 혹은 가우시안 분포라고 부름. 전체 면적은 1.

그래서 데이터를 통해 이 분포에 가장 가능성 높은 뮤와 시그마는 어떻게 찾을까?

여기에서는 우리가 과소평가하는 곳과 여기에서는 과대평가하는 곳이 있습니다. 그 차이를 더하거나 사실 그 제곱을 더해 맞는 적합의 정도 또는 나쁨의 척도를 얻을 수 있습니다. 그리고 우리는 모든 벡터 작업을 마친 후에 그것을 어떻게 수행하는지 자세히 살펴보고, 실제로 모든 미적분 작업을 마친 후에는 적합 매개변수인 시그마와 뮤를 변경함에 따라 적합의 정도가 어떻게 변화하는지 그래프로 나타낼 수 있습니다. 따라서 올바른 값, 여기에서 뮤에 대한 최적값, 그리고 여기에서 폭에 대한 최적값 시그마를 가지고 있다면 이렇게 그래프가 나타날 것입니다.

텍스트, 친필, 디자인이(가) 표시된 사진

자동 생성된 설명

우측 그래프: 뮤와 시그마에 대해 동일한 정도의 차이로 예측한 값들을 이은 것

ㄴ ~~이중에서 가장 적합한 값을 찾고 싶다?~~

인간의 얼굴, 사람, 의류, 친필이(가) 표시된 사진

자동 생성된 설명

뮤도 높게 예측, 변동성은 낮게 예측했을 때(변동성이 낮다 = 분포가 퍼져 있다)

뮤와 시그마를 변경해서 새로운 뮤에 대한 예측값들로 이동하여 결과를 확인할 수 있음.

이렇게 공간에서 이동하는 것은 벡터임. 벡터 공간에서 특정 지점으로의 이동하는 것.

그래서 벡터 공간을 이해하고 언덕을 내려가는 것을 이해하면 데이터 분포 예측하는 것이 가능함.

이때, 적합한 값의 곡선성을 이해하여 언덕을 내려가는 것이 미적분임. 즉, 미적분과 벡터를 이해하면 이러한 문제를 해결할 수 있음.

벡터는 공간을 축에 따라 방향성을 설명할 수 있음. 각 축은 구성요소로 벡터는 이러한 구성요소에 대한 공간을 설명할 수 있게 해줌.

**\*함수의 변수들에 대한 최적값을 찾을 때 공간으로 생각하는 것은 미친짓이 아님**

+아인슈타인은 상대성 이론을 고안할 때, 시간을 단순히 다른 차원으로 생각해서 ‘공간-시간’차원을 생각했음.

그는 우리가 사는 차원을 4차원으로, [ x y z t ]의 벡터로 표현하기도 했음.

**어쨋든 우리는 벡터 공간에서 가장 차이가 최소화되는 방향으로 최적화하려는 것.**

**이를 순서화하면 1. 함수를 나쁨(오차)을 나타낸 벡터 공간으로 표현**

1. **이 벡터에 대해 미적분을 수행**
   1. **=> 해당 등고선의 기울기를 구함**
   2. **=> 해당 벡터의 최소값을 찾음**

위 과정은 데이터와 머신러닝을 학습하는 과정.

What is a vector?

a. **A list of numbers**

Vectors are usually viewed by computers as an ordered list of numbers which they can perform "operations" on - some operations are very natural and, as we will see, very useful!

스크린샷, 텍스트, 도표, 디자인이(가) 표시된 사진

자동 생성된 설명

위 히스토그램의 각 구간을 벡터로 표현하면 아래와 같음.

텍스트, 폰트, 화이트, 영수증이(가) 표시된 사진

자동 생성된 설명

b. **Position in three dimensions of space and in one dimension of time.**

A vector in space-time can be described using 3 dimensions of space and 1 dimension of time according to some co-ordinate system.

c. **Something which moves in a space of fitting parameters.**

As we will see, vectors can be viewed as a list of numbers which describes some optimisation problem.

Vectors can be thought of in a variety of different ways - some geometrically, some algebraically, some numerically. In this way, there are a lot of techniques one can use to deal with vectors.

**Exploring parameter space**

잔차를 구하는 것은 모델이 데이터를 더 잘 예측하게 만들기 위해 필요한 지표 중 하나. 아래는 각 히스토그램 구간 별로 모델이 예측한 값(핑크)와 실제로 관측된 값(주황색)을 나타낸 그래프.

스크린샷, 텍스트, 도표, 디자인이(가) 표시된 사진

자동 생성된 설명

위 핑크 색 모델의 성능을 평가하기 위해 Sum of Squared Residuals, SSR와 같은 지표가 사용됨.

텍스트, 스크린샷, 도표, 디자인이(가) 표시된 사진

자동 생성된 설명

위 그래프는 뮤와 시그마 값을 수동으로 설정하여 SSR을 0.00051 이하로 맞춰본 그래프.

스크린샷, 원, 디자인이(가) 표시된 사진

자동 생성된 설명

위 그래프는 뮤와 시그마 값에 따른 SSR을 나타낸 그래프. 뮤와 시그마 값에 따라 SSR의 차이를 높이로 설정하고 등고선으로 동일한 SSR 값을 가지는 위치를 표시했음(뮤와 시그마 값에 따른 SSR의 분포를 쉽게 알아볼 수 있게).

스크린샷, 텍스트, 원이(가) 표시된 사진

자동 생성된 설명

위 그래프는 SSR을 높이로 설정한 전 그래프를 위에서 본 그래프. 등고선이 표시되어 있고 global minimum이 표시되어 있기 때문에 2차원으로도 파악가능.

머신러닝의 목표는 모델의 파라미터 값을 데이터에 가능한 잘 맞추는 것. 이것은 파라미터 공간에서 잔차의 값이 가장 작은 global minimum을 찾는 것임.

* The minimum of the surface is the best that the model can do, but it isn't a guarantee that the model itself is suitable, or that a better one doesn't exist!

하지만, 우리는 파라미터 공간 전체를 알 수 없기 때문에 가장 낮은 지점을 찾는 대신 더 낮은 지점이 어디인지 추측을 해야함.

이때 더 데이터에 맞추기 위해 어떤 변화를 줘야 하는지에 대한 Δp라는 다른 벡터를 정의할 수 있음.

예를 들어, p’ = p + Δp에서 p에서 Δp만큼 이동하여 더 데이터에 적합한 값을 찾기 위해 식을 작성할 수 있음(경사하강법)

스크린샷, 텍스트, 원이(가) 표시된 사진

자동 생성된 설명

Moving at right angles(수직) to contour lines(등고선) in the parameter space will have the greatest effect on the fit than moving in other directions.

등고선에 따라 움직이면 SSR값에 아무런 변화가 없음. However moving perpendicular(수직) to them can significantly improve or reduce the quality of the fit.

**Solving some simultaneous equations(선형 연립 방정식)**

1. **대체하기: 한 변수에 대해 정리하고 다른 식에 대입해서 변수의 값 찾기**
2. **제거하기: 다른 식을 연산해서 특정 변수를 제거할 수 있다면 제거해서 한 변수의 값을 먼저 찾기. 단, 다른 식의 변수 앞 계수가 동일하면 사용 불가**
3. **방정식에서 변수의 해를 알기 위해서는 변수 개수만큼 식이 필요함. 변수가 3개면 연립 방정식 3개.**

**Doing some vector operations(벡터 연산)**

스크린샷, 라인, 그래프, 디자인이(가) 표시된 사진

자동 생성된 설명

벡터 a는 피타고라스 정리를 통해 2sqrt(2)가 아니라.

텍스트, 폰트, 화이트, 번호이(가) 표시된 사진

자동 생성된 설명

이다. 즉, 격자선을 몇 칸씩 이동했는지로 표시한다.

스크린샷, 텍스트, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

벡터의 양의 스칼라 곱은 벡터 방향 그대로 스칼라 곱만큼 더 나아가면 됨. 혹은 각 요소별로 스칼라 곱을 연산.

스크린샷, 라인, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

음의 스칼라 곱을 계산하는 것은 벡터 방향의 반대 방향으로 스칼라 곱의 절댓값만큼 나아가면 됨. 혹은 각 요소별로 스칼라 곱을 연산.

스크린샷, 라인, 그래프, 폰트이(가) 표시된 사진

자동 생성된 설명

벡터를 더하는 것은 1) 각 벡터만큼 나아가서 구하기, 2) 각 벡터의 동일한 요소별로 합산하기가 있음.

결론적으로, **벡터연산은 1) 벡터의 개별 요소별로 연산, 2) 각 벡터 값대로 격자를 이동해서 최종 벡터 구하기의 두 가지 방법으로 가능**하다.

+ 선형 비선형 차이