Probability & Statistics

Department of AI Engineering

2024학년도 1학기

Prof. Suk-Hwan Lee

Artificial Intelligence

Creating the Future

Dong-A University

Division of Computer Engineering & Artificial Intelligence

Probability & Statistics

<table>
<tr><td rowspan="2">수업소개</td><td>교과목의
필요성</td><td colspan="3">인공지능 전문가로서 머신러닝/딥러닝의 핵심기초이론인 확률변수와 확률분포, 평균과분산 등 통계 내용을 기반으로 머신러닝/딥러닝 모델의 이론과, 모델의 설계, 구현, 평가의 전 과정을 이해하기 위하여 확률 및 통계를 학습하는 것이 필수이다.</td></tr>
<tr><td>교과목
개요</td><td colspan="3">본 강의는 머신러닝/딥러닝 모델의 핵심 개념인 확률론의 기초 이론, 확률의 개념, 확률 분포 개념과 머신러닝/딥러닝에 사용되는 핵심 확률 이론 등을 학습한다. 본 강의를 통하여 머신러닝/딥러닝 모델에 사용되는 확률 이론을 소개하고자 한다.</td></tr>
<tr><td>수업목표</td><td colspan="4">본 강의는 머신러닝/딥러닝 모델의 핵심 개념인 확률론의 기초 이론인 1) 확률 정의, 조건부 확률, 베이지 이론, 독립, 랜덤변수 및 기댓값 등의 주요 확률 개념과 2) 분산, 베르누이 및 이항 분포, 이산 분포, 결합 분포, 조건 분포 등의 확률 분포 개념과 3) 공분산, 상관분석, 중심극한정리 등의 확률 심화 개념과 4) 머신러닝을 위한 최대우도비, 최</td></tr>
<tr><td>사전학습</td><td colspan="4">1. 고등학교에서 확률및통계 관련 교과목을 이수하지 않은 학생들은 어렵게 느낄 수 있으나, 대학기초수학, 또는 선형대수학 등의 사전 이수한 경우, 본 강의를 이해하는데 도움이 된다.
2. 인공지능을 위한 확률및통계 교재가 많지 않은 관계로 확률과 관련된 자료는 별도로 소개할 예정이다.</td></tr>
<tr><td rowspan="3">교재 및
참고자료</td><td>주교재</td><td colspan="3">별도 강의자료 제공 예정 (LMS 수업자료실 활용)</td></tr>
<tr><td>참고자료</td><td colspan="3">1. 인공지능, 머신러닝, 딥러닝과 관련된 자료들 (Textbook, Blog 등)
2. 데이터 과학을 위한 통계 2판, Oreilly 한빛미디어</td></tr>
<tr><td>참고사이트</td><td colspan="3">http://web.stanford.edu/class/archive/cs/cs109/cs109.1166/handouts/overview.html</td></tr>
<tr><td rowspan="2">수업 방법</td><td>강의식</td><td>실험/실습</td><td></td><td></td></tr>
<tr><td colspan="4">1. 대면 강의로 진행함.
2. 강의시간에는 대부분 이론 중심으로 진행하고, 경우에 따라 실습(파이썬, pytorch 등)을 할 수 있음.
※ 주별 수업계획은 일정 또는 강의 진도에 따라 학습목표 및 학습내용이 변경될 수 있음</td></tr>
</table>

Probability & Statistics

<table>
<tr><td rowspan="3">학습 평가 방법</td><td>종류</td><td>출석</td><td>과제</td><td>임의평가</td><td>중간시험</td><td>기말시험</td><td>기타</td><td>합계</td></tr>
<tr><td>비율</td><td>15</td><td>5</td><td>0</td><td>40</td><td>40</td><td>0</td><td>100%</td></tr>
<tr><td>방법</td><td colspan="7">1. 출석 : 결석 1회 감점 1점(결석 한 주의 온라인 학습결과 제출 시 출석 인정), 지각 3회 감점 1점

2. 시험 : 중간시험/기말시험, 문제 난이도에 따라 객관식/서술식/논술형 등 유형으로 문제 제시

3. 과제 : 주요 Chapter에 대하여 미니 과제 제시 예정임. (강의 일정에 따라 과제 없을 수도 있음)

4. 기타 : 수업태도 불성실한 학생에게 감점이 있음 (강의 중에 이동하거나, 방해되는 행위 출석 감점)
=> 청강자의 매너, 이메일 매너 준수하기

※ 강의 일정과 상황에 따라 2,3번은 변경될 수 있으며, 사전에 공지할 예정임.

※ SW중심대학사업 및 학부 관련 프로그램 참여하기 (프로그램 참여도에 따라 가점 부여할 수 있음)</td></tr>
</table>

Introduction

Main Contents

- **https://web.stanford.edu/class/cs109/**

Materials

- **https://chrispiech.github.io/probabilityForComputerScientists/en/index.html**

Tensorflow Probability

- **https://www.tensorflow.org/probability?hl=ko**
- **https://github.com/tensorflow/probability**

Introduction

Main Contents

- **https://web.stanford.edu/class/cs109/**

Lecture 1: Introductions	slides annotated video concept-check		Piech: Counting
Lecture 2: Permutations and Combinations	slides annotated video concept-check just-for-fun	PSet 1 Out	Piech: Combinatorics
Lecture 3: Axioms of Probability	slides annotated video demo concept-check just-for-fun		Ross: Ch 2.1-2.5, 2.7 Piech: Probability, Equally Likely Outcomes
Observance of Martin Luther King Day. No lecture.			
Lecture 4: Conditional Probability and Bayes Rule	slides annotated video demo concept-check just-for-fun		Ross: Ch 3-1-3.3 Piech: Conditional Probability, Law of Total Probability, Bayes Theorem
Section 1: Combinatorics and Probability	handout solution		
Lecture 5: Independence	slides annotated video concept-check	PSet 1 In, PSet 2 Out	Ross: Ch 3.4-3.5 Piech: Independence

Introduction

Main Contents

- **https://web.stanford.edu/class/cs109/**

Lecture 6: Random Variables and Expectation	slides annotated video concept-check		Ross: Ch 4.1-4.4 Piech: Random Variables, Probability Mass Functions
Lecture 7: Variance, Bernoulli, Binomial	slides annotated video demo concept-check just-for-fun		Ross: Ch 4.5-4.6 Piech: Variance, Bernoulli, Binomial
Section 2: Random Variables and Expectation	handout solution		
Lecture 8: Poisson and Approximations	slides annotated video concept-check		Ross: 4.7-4.10 Piech: Poisson
Lecture 9: Continuous Random Variables	slides annotated video concept-check	PSet 2 In	Ross: Ch 5.1-5.3, 5.5 Piech: Continuous RVs
Lecture 10: The Normal Distribution	slides annotated video demo concept-check	PSet 3 Out	Ross: Ch 5.4, Piech: Gaussian Distributions, Binomial Approximations
Section 3: Discrete and Continuous Random Variables	handout solution		

Introduction

Main Contents

- **https://web.stanford.edu/class/cs109/**

Lecture 11: Joint Distributions	slides annotated video concept-check		Ross: Ch 6.1, Piech: Joint Probability, Multinomials, Federalist Papers
Lecture 12: Independent Random Variables	slides annotated video concept-check		Ross: Ch 6.2-6.3, Piech: Adding Random Variables
Lecture 13: Joint RV Statistics	slides annotated video concept-check just-for-fun		Ross: Ch 6.4-6.5, Piech: Correlation
Section 4: Normal Distributions and Joint Distributions	handout solution		
Lecture 14: Conditional Expectation	slides annotated video concept-check just-for-fun	PSet 3 In, PSet 4 Out, CS109 Challenge Out	Ross: Ch 7.1-7.2, Piech: *No assigned reading*
Lecture 15: General Inference	slides annotated video concept-check demo		Ross: *No assigned reading,* Piech: Inference
Lecture 16: Continuous Joint Distributions	slides annotated video concept-check		Ross: Ch 6.1, Piech: Continuous Joint Distributions

Introduction

Main Contents

- **https://web.stanford.edu/class/cs109/**

Section 5: Conditional Expectation	handout solution		
Lecture 17: Continuous Joint Distributions II	slides video concept-check		Ross: Ch 7.3-7.4, Piech: *no assigned reading*
Lecture 18: Central Limit Theorem	slides annotated video concept-check sampling-demo clt-demo		Ross: Ch 8.3, Piech: Central Limit Theorem
Section 6: Continuous Joint Distributions, Central Limit Theorem	handout solution		
Lecture 19: Sampling and Bootstrapping	slides annotated video demo concept-check just-for-fun	PSet 5 Out	Ross: *No assigned reading*, Piech: Sampling, Bootstrapping
Lecture 20: Parameters and MLE	slides annotated video demo concept-check		Ross: *No assigned reading*, Piech: MLE, MLE Normal, MLE Mixture

Introduction

Main Contents

- **https://web.stanford.edu/class/cs109/**

Lecture 21: Beta	slides annotated video concept-check just-for-fun		Ross: Ch 5.6.1-5.6.4, 7.5-7.6, Piech: Beta
Section 7: Boostrapping and MLE	handout solution colab		
Lecture 22: Maximum a Posteriori	slides annotated concept-check		Ross: *No assigned reading.*, Piech: Maximum a Posteriori
Lecture 23: Naive Bayes		PSet 5 In, PSet 6 Out	Ross: *No assigned reading.*, Piech: Machine Learning, Naive Bayes
Lecture 24: Linear Regression, Gradient Ascent			Ross: *No assigned reading.*, Piech: *No assigned reading.*
Section 8: Parameter Estimation and Naive Bayes			

Introduction

Main Contents

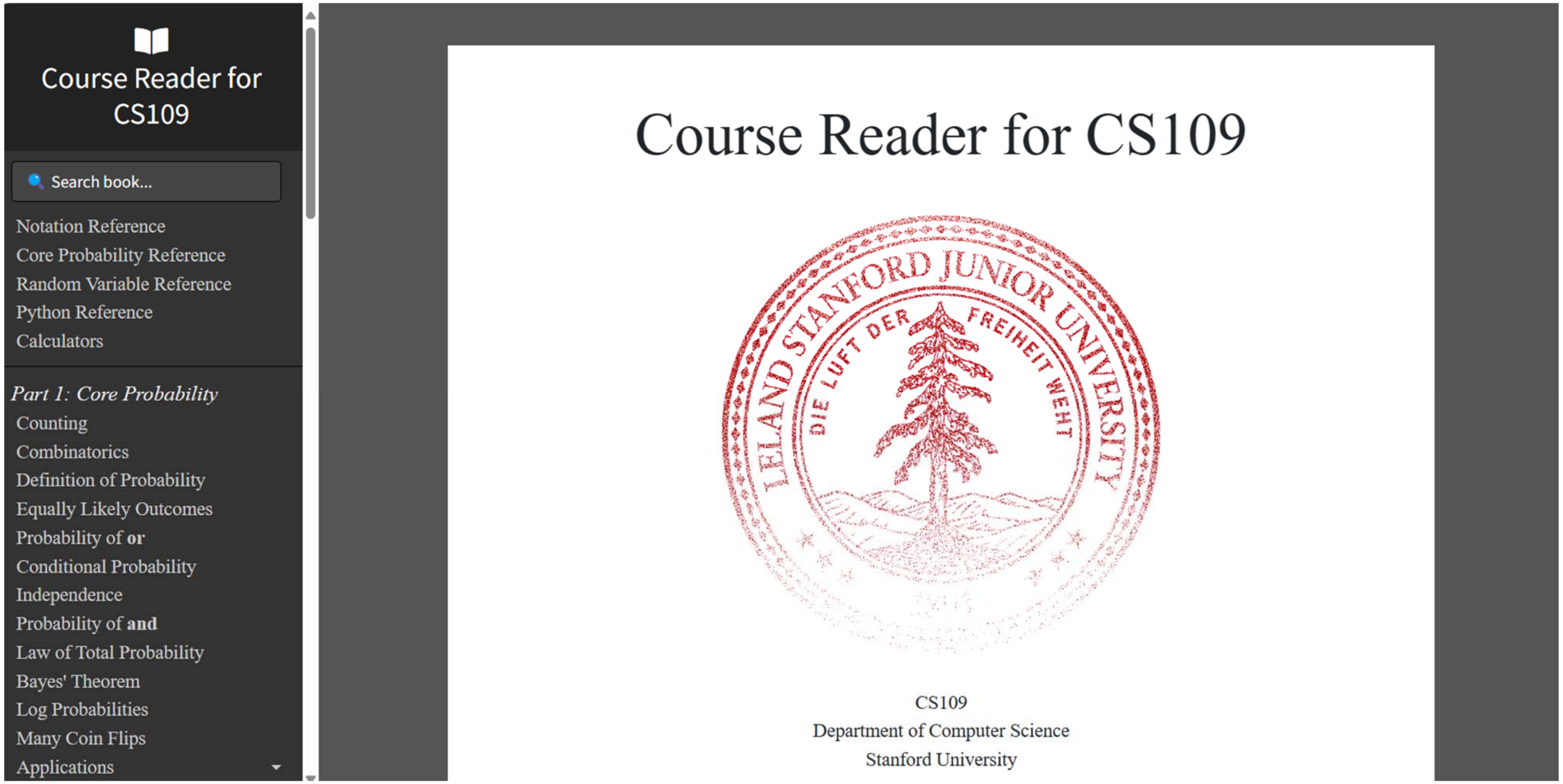
- **https://web.stanford.edu/class/cs109/**

Lecture 25: Logistic Regression		Ross: *No assigned reading.*, Piech: Logistic Regression
Lecture 26: Deep Learning		Ross: *No assigned reading.*, Piech: *No assigned reading.*
Lecture 27: Ethics in Probability and AI	PSet 6 In	Ross: *No assigned reading.*, Piech: *No assigned reading.*
Section 9: Final Exam Review		

Introduction

Materials

- **https://chrispiech.github.io/probabilityForComputerScientists/en/index.html**



Notation References

Core Probability

Notation	Meaning
E	Capital letters can denote events
A	Sometimes they denote sets
$\|E\|$	Size of an event or set
E^C	Complement of an event or set
EF	And of events (aka intersection)
E and F	And of events (aka intersection)
$E \cap F$	And of events (aka intersection)
E or F	Or of events (aka union)
$E \cup F$	Or of events (aka union)
$\text{count}(E)$	The number of times that E occurs
$\mathrm{P}(E)$	The probability of an event E
$\mathrm{P}(E\|F)$	The conditional probability of an event E given F
$\mathrm{P}(E, F)$	The probability of event E and F
$\mathrm{P}(E\|F, G)$	The conditional probability of an event E given both F and G
$n!$	n factorial
$\binom{n}{k}$	Binomial coefficient
$\binom{n}{r_1, r_2, r_3}$	Multinomial coefficient

Notation References

Random Variables

Notation	Meaning
x	Lower case letters denote regular variables
X	Capital letters are used to denote random variables
K	Capital K is reserved for constants
$\mathrm{E}[X]$	Expectation of X
$\mathrm{Var}(X)$	Variance of X
$\mathrm{P}(X=x)$	Probability mass function (PMF) of X, evaluated at x
$\mathrm{P}(x)$	Probability mass function (PMF) of X, evaluated at x
$f(X=x)$	Probability density function (PDF) of X, evaluated at x
$f(x)$	Probability density function (PDF) of X, evaluated at x
$f(X=x, Y=y)$	Joint probability density
$f(X=x \mid Y=y)$	Conditional probability density
$F_X(x)$ or $F(x)$	Cumulative distribution function (CDF) of X
IID	Independent and Identically Distributed

Notation References

Parametric Distributions

Notation	Meaning
$X \sim \mathrm{Bern}(p)$	X is a Bernoulli random variable
$X \sim \mathrm{Bin}(n, p)$	X is a Binomial random variable
$X \sim \mathrm{Poi}(p)$	X is a Poisson random variable
$X \sim \mathrm{Geo}(p)$	X is a Geometric random variable
$X \sim \mathrm{NegBin}(r, p)$	X is a Negative Binomial random variable
$X \sim \mathrm{Uni}(a, b)$	X is a Uniform random variable
$X \sim \mathrm{Exp}(\lambda)$	X is a Exponential random variable
$X \sim \mathrm{Beta}(a, b)$	X is a Beta random variable

Introduction

Tensorflow Probability

- **https://www.tensorflow.org/probability?hl=ko**

TensorFlow Probability는 확률적 추론 및 통계 분석을 위한 라이브러리입니다

TensorFlow Probability(TFP)는 최신 하드웨어(TPU, GPU)에서 확률 모델 및 딥 러닝을 손쉽게 결합할 수 있도록 도와주는 TensorFlow 기반의 Python 라이브러리입니다. TFP는 데이터를 이해하고 예측하기 위해 도메인 지식을 인코딩하려는 데이터 과학자, 통계 전문가, ML 연구자 및 실무자용 라이브러리입니다. TFP에는 다음 항목이 포함됩니다.

- 다양한 확률 분포 및 바이젝터
- 확률 레이어 및 `JointDistribution` 추상화를 비롯한 딥 확률 모델을 빌드하는 도구
- 변이 추론 및 마르코프 체인 몬테카를로법
- Nelder-Mead, BFGS 및 SGLD와 같은 옵티마이저

TFP는 TensorFlow의 이점을 계승하므로 모델 탐색 및 생산의 수명 주기 전반에 걸쳐 단일 언어를 사용하여 모델을 빌드, 조정 및 배포할 수 있습니다. TFP는 오픈소스이며 GitHub에서 사용 가능합니다. 시작하려면 TensorFlow Probability 가이드를 참조하세요.

```
import tensorflow as tf
import tensorflow_probability as tfp

# Pretend to load synthetic data set.
features = tfp.distributions.Normal(loc=0., scale=1.).sam
labels = tfp.distributions.Bernoulli(logits=1.618 * featu

# Specify model.
model = tfp.glm.Bernoulli()

# Fit model given data.
coeffs, linear_response, is_converged, num_iter = tfp.glm
    model_matrix=features[:, tf.newaxis],
    response=tf.cast(labels, dtype=tf.float32),
    model=model)
# ==> coeffs is approximately [1.618] (We're golden!)
```

CO 메모장에서 실행

Introduction

Pytorch Distributions

- **https://pytorch.org/docs/stable/distributions.html**

PROBABILITY DISTRIBUTIONS - TORCH.DISTRIBUTIONS

The `distributions` package contains parameterizable probability distributions and sampling functions. This allows the construction of stochastic computation graphs and stochastic gradient estimators for optimization. This package generally follows the design of the TensorFlow Distributions package.

It is not possible to directly backpropagate through random samples. However, there are two main methods for creating surrogate functions that can be backpropagated through. These are the score function estimator/likelihood ratio estimator/REINFORCE and the pathwise derivative estimator. REINFORCE is commonly seen as the basis for policy gradient methods in reinforcement learning, and the pathwise derivative estimator is commonly seen in the reparameterization trick in variational autoencoders. Whilst the score function only requires the value of samples $f(x)$, the pathwise derivative requires the derivative $f'(x)$. The next sections discuss these two in a reinforcement learning example. For more details see Gradient Estimation Using Stochastic Computation Graphs .

Score function

When the probability density function is differentiable with respect to its parameters, we only need `sample()` and `log_prob()` to implement REINFORCE:

$$\Delta\theta = \alpha r \frac{\partial \log p(a|\pi^\theta(s))}{\partial\theta}$$

Probability distributions - torch.distributions
Score function
Pathwise derivative
+ Distribution
+ ExponentialFamily
+ Bernoulli
+ Beta
+ Binomial
+ Categorical
+ Cauchy
+ Chi2
+ ContinuousBernoulli
+ Dirichlet
+ Exponential
+ FisherSnedecor
+ Gamma
+ Geometric
+ Gumbel
+ HalfCauchy
+ HalfNormal
+ InverseGamma
+ Kumaraswamy
+ LKJCholesky
+ Laplace
+ LogNormal
+ LowRankMultivariateNormal
+ MixtureSameFamily
+ Multinomial
+ MultivariateNormal
+ NegativeBinomial
+ Normal
+ OneHotCategorical
+ Pareto
+ Poisson
+ RelaxedBernoulli
+ LogitRelaxedBernoulli
+ RelaxedOneHotCategorical
+ StudentT
+ TransformedDistribution
+ Uniform
+ VonMises
+ Weibull
+ Wishart
+ *KL Divergence*
+ *Transforms*
+ *Constraints*
+ *Constraint Registry*