第四讲
概率论选讲
Lecture 4
Selected topics of Probability

明玉瑞 Yurui Ming

yrming@gmail.com

声明
Disclaimer

- 本讲义在准备过程中由于时间所限，所用材料来源并未规范标示引用来源。所引材料仅用于教学所用，作者无意侵犯原著者之知识版权，所引材料之知识版权均归原著者所有；若原著者介意之，请联系作者更正及删除。

  The time limit during the preparation of these slides incurs the situation that not all the sources of the used materials (texts or images) are properly referenced or clearly manifested. However, all materials in these slides are solely for teaching and the author is with no intention to infringe the copyright bestowed on the original authors or manufacturers. All credits go to corresponding IP holders. Please address the author for any concern for remedy including deletion.

概率
Probability

- 概率是对某个事件发生的可能性的度量，该值表示期望发生的情况与所有可能发生的情况的总数的比率。

  Probability is the measure of how likely it is some event will occur, a number that expresses the ratio of favorable cases to the whole number of cases possible.

- 例如，当掷硬币时，正面朝上的几率是 50%。这是因为一枚硬币有两个面，所以硬币正面朝上的概率为 50%，反面朝上的概率为 50%。

  For example, the chance of landing on heads is 50% when toss a coin. This is because a coin has two sides, so there is a 50% chance that the coin will land on heads and 50% that a coin will land on tails.

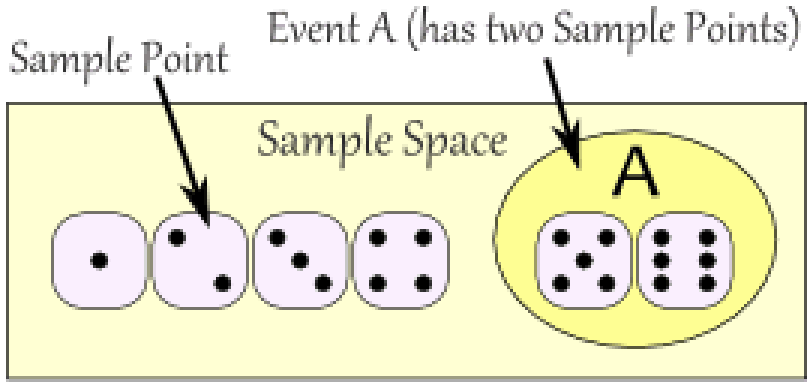
概率
Probability

- 一些概念，如试验，采样空间，事件

  Terminology, for example, trial, sampling space, event

- 随机变量是在样本空间上定义的函数，进一步说，随机变量是为样本空间中的每个结果分配数值的规则。

  A random variable is a function defined on a sample space. Further, a random variable is a rule that assigns a numerical value to each outcome in a sample space.



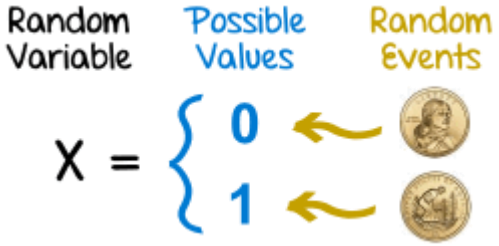
概率
Probability

- 随机变量本身的取值并不是重点，重点是取特定值的可能性的数值度量，即概率

  The key is not the rule of assigning values to the variable itself, it is the values measure the values upon experiment.

- 随机变量分为离散型随机变量与连续型随机变量

  Random variables are categorized into discreate random variable and continuous random variable.



$$P(X=0)=P(X=1)=\frac{1}{2}$$

概率
Probability

- 对于概率分布，我们先考虑离散变量的情况。假设X为一离散随机变量，其可取值为x_1，x_2，x_3，…。假设X取x_k的概率为$P(X = x_k) = f(x_k)$，则一般地，我们称满足如下条件的$f(x) \stackrel{\text{def}}{=} P(X = x)$为概率函数或概率分布：

$$f(x) \geq 0$$

$$\sum_x f(x) = 1$$

We consider first the probability distribution for a discreate random variable. Let X be a discrete random variable and suppose that the possible values that it can assume are given by x_1, x_2, x_3, ….. Suppose also that these values are assumed with probabilities given by $P(X = x_k) = f(x_k)$, where $k = 1, 2, 3, \cdots$. It is convenient to introduce the probability function, also referred to as probability distribution, given by $f(x) \stackrel{\text{def}}{=} P(X = x)$ on condition that the following two criteria satisfied:

$$f(x) \geq 0$$

$$\sum_x f(x) = 1$$

概率
Probability

- 对于连续型随机变量X，同样可以定义其概率分布$f(x) \overset{\text{def}}{=} P(X = x)$，此时要求$\int f(x)dx = 1$。

  Probability distribution also holds for a continuous random variable X given by $f(x) \overset{\text{def}}{=} P(X = x)$ on condition that $\int f(x)dx = 1$.

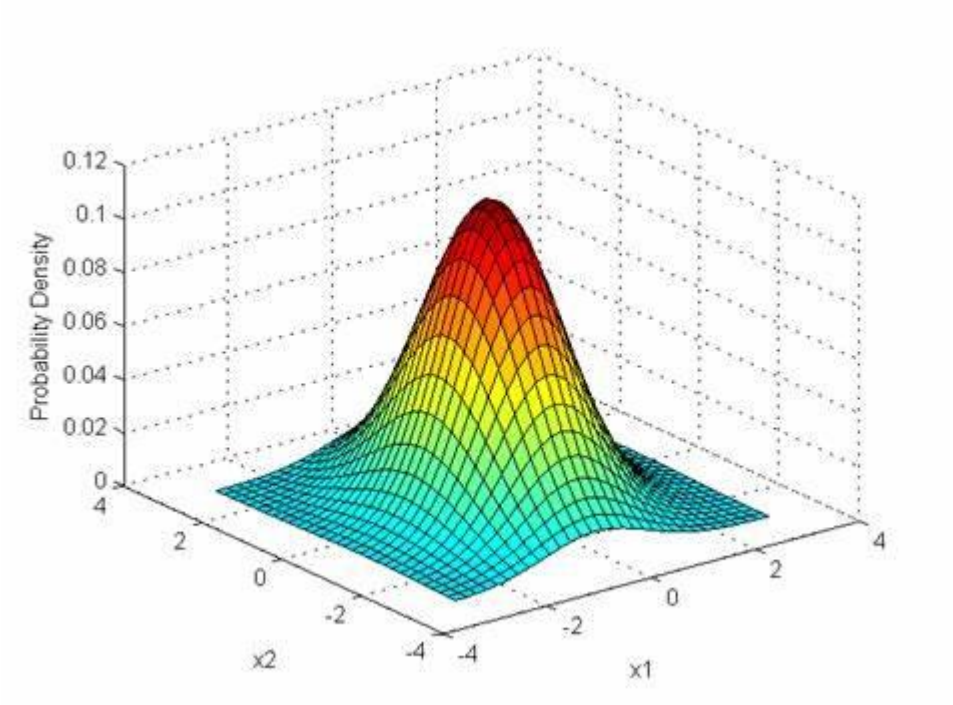
- 简而言之，概率衡量了一个事件出现的可能性；概率分布刻画了所有可能出现的事件与其出现可能性大小的关系

  Probability calculates the likelihood of one event; Probability distribution depicts the relationship between each possible outcome for a random variable and their probabilities.

概率
Probability

- 多元随机变量的分布

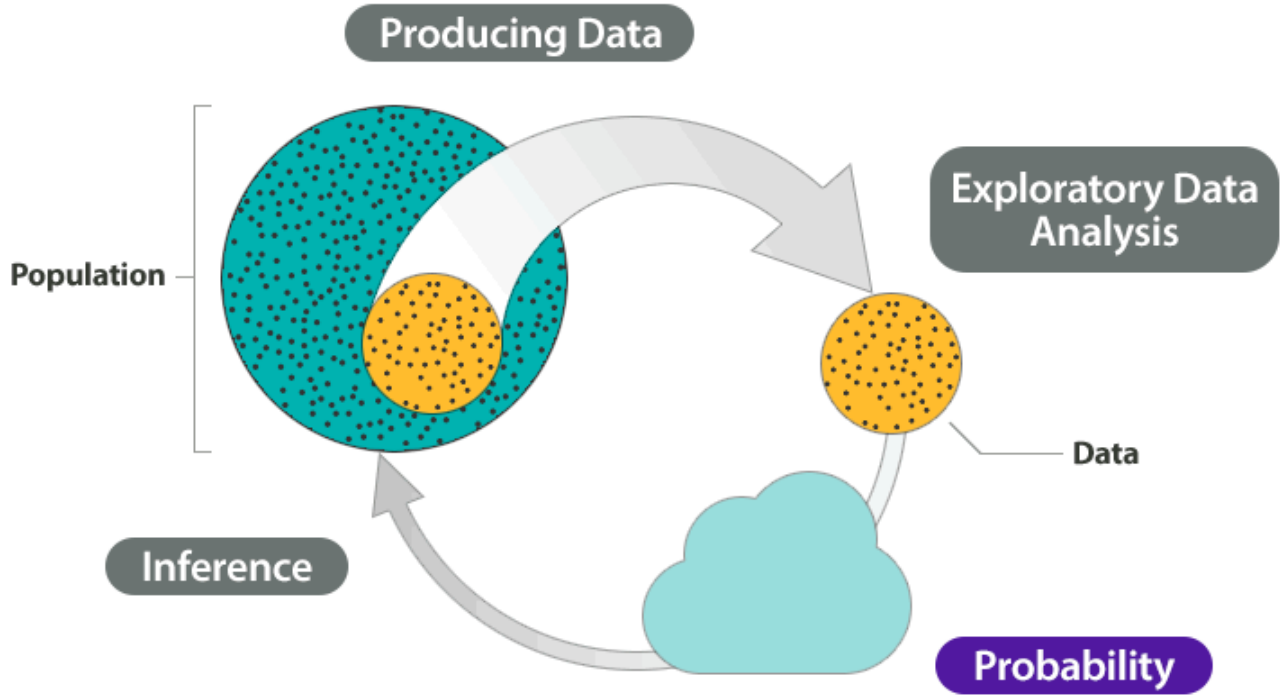
  Distribution of multi-random variables



概率
Probability

- 如何求得概率分布？为何求得概率分布？

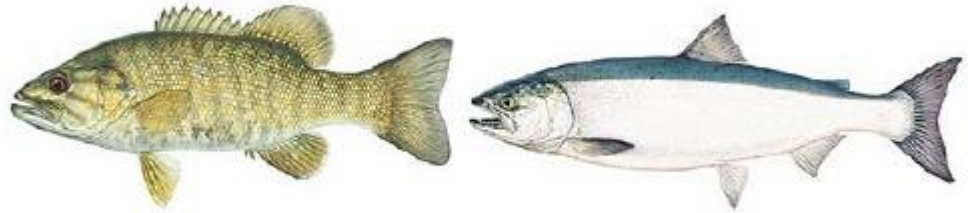
  How and why to derive probability distribution?



贝叶斯决策理论
Bayesian Decision Theory

- 例子：如何判定鲈鱼与三文鱼。一捕鱼公司从近海捕捞一批鱼类，在产品线上做分拣。如何自动判定从传送出口送上来的是鲈鱼，还是三文鱼？

  How to classify see bass and salmon. A fishing company wants to sort the fishes caught in the near see on the assembly. How to decide a fish delivered from the conveyer is see bass or salmon?



(a) (b)

贝叶斯决策理论
Bayesian Decision Theory

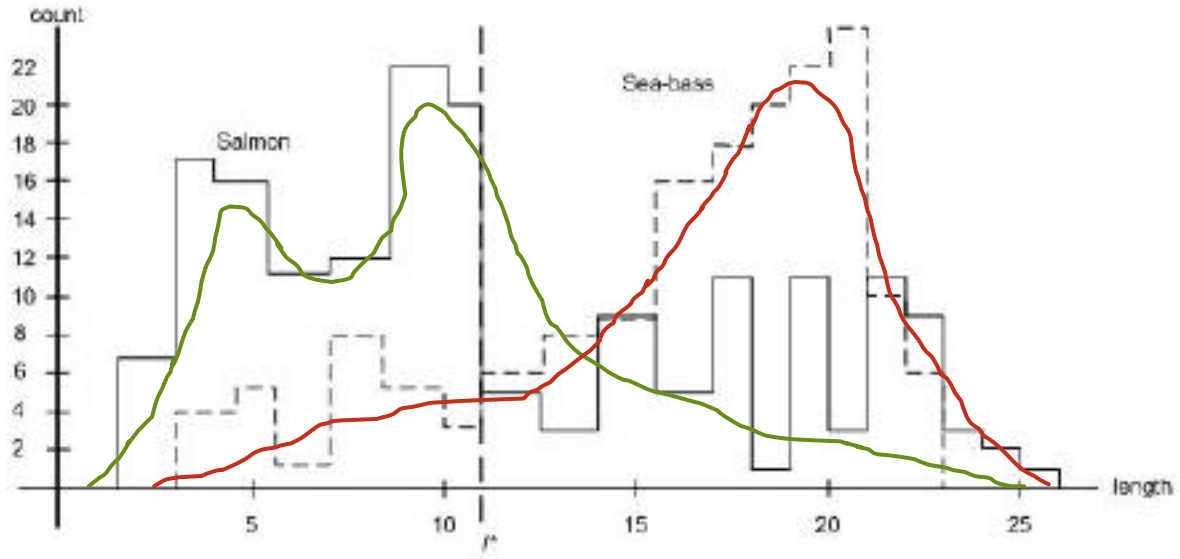
- 如果没有任何信息，则随机猜测，如果有对鲈鱼或三文鱼的先验认识，则根据这个这个认识做判断。例如，令X表示为何种鱼的随机变量，$X=0$表示鲈鱼，$X=1$表示三文鱼，如果根据先验知识，有$P(X=0)=0.25$，$P(X=1)=0.75$，则无论传送带传送上来什么鱼，统一猜三文鱼。为什么？

  If there is no any other information, the only way to perform is a random guess. But, if there exists some priori knowledge about see bass and salmon, the decision can take this priori into consideration. For example, let X be the random variable representing the kind of fish, and $X=0$ indicates the fish currently conveyed is see bass, while $X=0$ indicates salmon. If according to experience or survey, we have $P(X=0)=0.25$, $P(X=1)=0.75$, then a universal strategy is to always guess salmon. Why?

贝叶斯决策理论
Bayesian Decision Theory

- 假设在长期的实践中，观察到鱼体长度与种类的关系。经平滑，得到以下概率分布函数。则又该如何判断呢？

  Suppose during the practice, the relation between the class of fish and its corresponding length has been drawn. The probability distributions of the length of fishes are shown below after smoothing. Then how to decide?



贝叶斯决策理论
Bayesian Decision Theory

- 在接着讲述之前，先简单地介绍一下条件概率。概率$P(B|A)$描述的是事件B在另一个事件A已经发生的条件下的概率，这也正是这个名称的由来。这里A和B可能是相互独立的两个事件（即A和B的发生互不相关），也可能不是。

  Before continuing the example, conditional probability is briefly introduced. $P(B|A)$ depicts the probability of event B occurs given the happening of event A, and therefore the name conditional probability comes. Here A and B can be two independent events (aka, occurrence of event A impact nothing on event B and vice versa), or not.

- A和B同时发生的概率记为$P(A \cap B)$或$P(AB)$；如果A和B是相互独立的两个事件，则$P(A \cap B) = P(A) \times P(B)$。

  The probability of the events A and B occur simultaneously is denoted by $P(A \cap B)$ or $P(AB)$. If A and B are independent from each other, then it holds $P(A \cap B) = P(A) \times P(B)$

贝叶斯决策理论
Bayesian Decision Theory

- 根据条件概率的定义，不难知

  It is easily drawn from the definition of conditional probability that

$$P(B|A) = \frac{P(AB)}{P(A)}$$

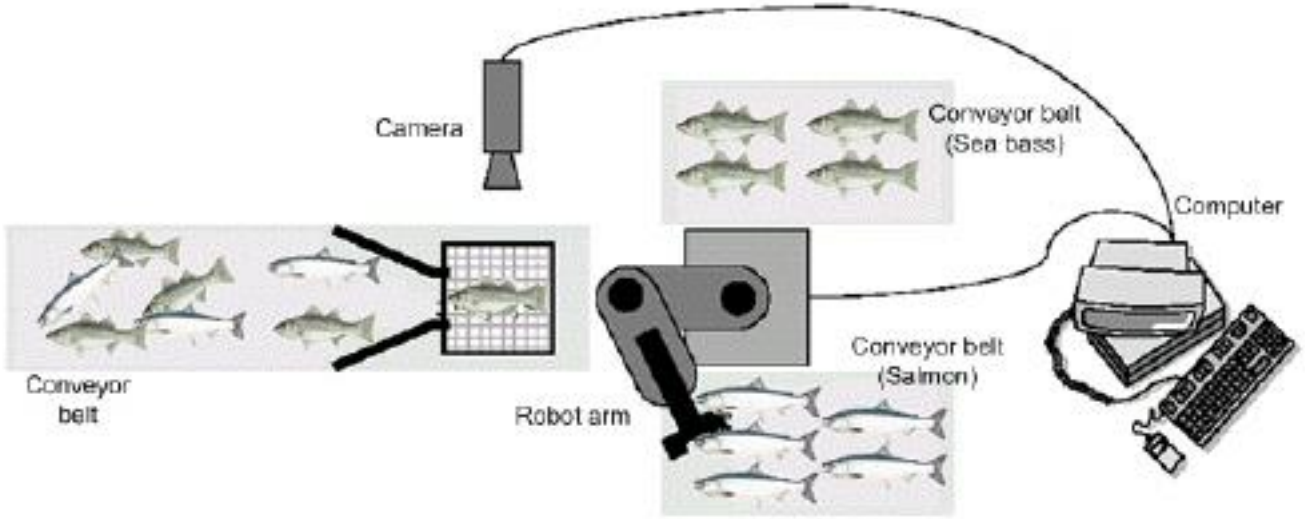
- 针对判断鱼的种类的例子，沿用符号，即X表示为何种鱼的随机变量，L表示为鱼的长度的随机变量。则$P(L|X=0)$表示鲈鱼的身长取某个值的概率，其光滑过后的概率分布函数可记为$p(L|X=0)$，$P(L|X=1)$表示三文鱼的身长取某个值的概率，其光滑过后的概率分布函数可记为$p(L|X=1)$。

  Taken the fish classification example, let X denote the random variable of fish class, and L for the length of the fish. Then $P(L|X=0)$ indicates the probability of a specific value of body length for sea bass, and $p(L|X=0)$ denotes the probability distribution function after smoothing $P(L|X=0)$. The same deduction applies to the case of salmon.

贝叶斯决策理论
Bayesian Decision Theory

- 现在有了鱼的体长的概率分布函数。则对传送上来的鱼，通过设备取得鱼的体长后，又该如何判断呢？

  Now we have the distribution functions of the fish length of different kinds. Suppose via apparatus we obtained the length of the fish. Then how to decide subsequently?



贝叶斯决策理论
Bayesian Decision Theory

- 我们知道$p(L|X=0)$和$p(L|X=0)$，若同时有$P(X=0)=0.01$，$P(X=1)=0.99$，则如何判断呢？若$P(X=0)=0.3$，$P(X=0)=0.7$呢？

  We already know $p(L|X=0)$ and $p(L|X=0)$. And if further we have $P(X=0)=0.01$，$P(X=1)=0.99$, Then how we make the decision? If $P(X=0)=0.3$ and $P(X=0)=0.7$?

- 考虑事件A，称$\bar{A}$为事件A的补集，若$P(A)$表示出现事件A的概率，则$P(\bar{A})$表示出现事件A的概率，显然$P(A)+P(\bar{A})=1$

  For event A, $\bar{A}$ is call the complement of A. If $P(A)$ denotes the probability of the occurrence of event A, then $P(\bar{A})$ denotes the opposite. Obviously, we have $P(A)+P(\bar{A})=1$.

贝叶斯决策理论
Bayesian Decision Theory

- 直观上，我们有以下式子成立：$P(A) = P(A \cap 1) = P\big(A \cap (B \cup \bar{B})\big) = P\big((A \cap B) \cup (A \cap \bar{B})\big) = P(A \cap B) + P(A \cap \bar{B})$。即一件事情发生的概率，等于另一件事情发生时，该事件发生的概率，加上另一件事情不发生时，该事件发生的概率。

  Intuitively, we have $P(A) = P(A \cap 1) = P\big(A \cap (B \cup \bar{B})\big) = P\big((A \cap B) \cup (A \cap \bar{B})\big) = P(A \cap B) + P(A \cap \bar{B})$ holds. It means, the probability of one event occurs, equals to the probability of the occurrence of this event on condition that another event occurs, plus that of the case the same another event doesn't occur.

- 注意到$P(AB) = P(B)P(A|B)$，则$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$，称为全概率公式。

  Notice $P(AB) = P(B)P(A|B)$, 则$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$, named law of total probability.

贝叶斯决策理论
Bayesian Decision Theory

- 贝叶斯定理：

  Bayes Theorem

  $$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

  其中$P(A)$为先验概率，$P(B|A)$为条件概率，$P(A|B)$为后验概率，$P(AB)$为联合概率，亦写成$P(A \cap B)$或$P(A,B)$。

  Here $P(A)$ is called prior probability, $P(B|A)$ is called conditional probability, $P(A|B)$ is called posterior probability, $P(AB)$ is called joint probability, sometimes also denoted as $P(A \cap B)$ or $P(A,B)$.

- 实际中，$P(B)$往往会用全概率公式展开，但是在一些方法比如朴素贝叶斯的应用中，会被抵消掉。

  In practice, $P(B)$ will be expanded using law of total probability, however, in some applications such as naïve Bayesian method, $P(B)$ will be cancelled out.

贝叶斯决策理论
Bayesian Decision Theory

- 若已知鱼体长度的情况下，求鱼为鲈鱼的概率：

  Given the length l, calculate the probability of sea bass:

$$P(X=0|L=l)=\frac{P(X=0\cap L=l)}{P(L=l)}=\frac{P(L=l|X=0)P(X=0)}{P(L=l)}$$

- 若已知鱼体长度的情况下，求鱼为三文鱼的概率：

  Given the length l, calculate the probability of salmon:

$$P(X=1|L=l)=\frac{P(X=1\cap L=l)}{P(L=l)}=\frac{P(L=l|X=1)P(X=1)}{P(L=l)}$$

  可以通过比较后验概率的大小判断，定义如下$D(X|L=l)$：

  The decision can be made based on the posterior probability. Define $D(X|L=l)$ as follows:

$$D(X|L=l)=\frac{P(X=0|L=l)}{P(X=1|L=l)}=\frac{P(L=l|X=0)P(X=0)}{P(L=l|X=1)P(X=1)}$$

贝叶斯决策理论
Bayesian Decision Theory

- 若$D(X|L=l)>1$，则判断为鲈鱼，否则判断为三文鱼。

  If $D(X|L=l)>1$, Then the output is sea bass, otherwise it is salmon.

- 注意到我们对$P(L|X=0)$或$P(L|X=1)$，均是得到平滑之后的概率分布函数，即$p(L|X=0)$或$p(L|X=1)$，但对$D(X|L=l)$的定义，在有的项是概率分布，有的项是概率分布函数的时候，依然有效，即依然可以依据以下定义的$D(X|L=l)$进行判断：

  It is mentioned here that for $P(L|X=0)$ or $P(L|X=1)$, what we obtained are the distribution functions after smoothing, aka, $p(L|X=0)$ and $p(L|X=1)$. However, in practice, the definition of $D(X|L=l)$ still holds by mixing probability and probability distribution together, just as above. We can still rely on the value of $D(X|L=l)$ to make the decision.

$$D(X|L=l)=\frac{p(L=l|X=0)P(X=0)}{p(L=l|X=1)P(X=1)}$$

神经网络的概率解释
Interpretation of Neural network from the probabilistic perspective

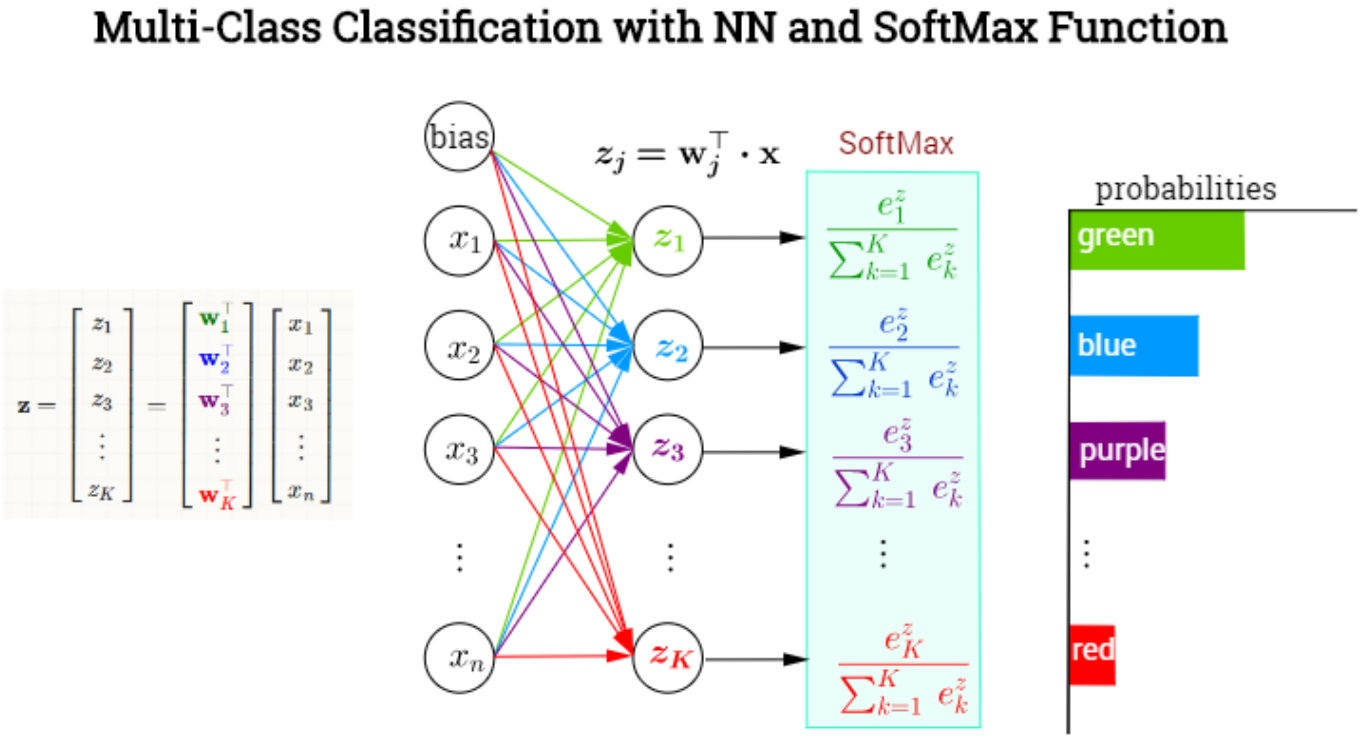
- 考虑用神经网络进行猫和狗的概率分类问题。X表示图片数据，y表示图片辨识结果（标签），均为随机变量；令$f(y|X;w)$表示神经网络，同时看作一个参数化的概率模型（依赖于w），即观察到某个图片输入时，输出为特定类别的概率。令$P(y=0)$表示是猫的图片概率，$P(y=1)$表示是狗的图片的概率；$P(X|y=0;w)$表示当标签是猫时，图片的特征的概率分布；$P(X|y=1;w)$表示当标签是狗时，图片的特征的概率分布；令$P(y=0|X=x;w)$表示当观察到输入为x时，输出是猫的概率，$P(y=1|X=x;w)$表示当观察到输入为x时，输出是狗的概率。

  Consider the application of neural networks for image classification. Let X and y be random variables to denote image data and image prediction (label) respectively. Let $f(y|X;w)$ denote the network, which is treated as a parameterized probabilistic model (by parameter w) simultaneously. It is interpreted as for a given image data observed, the probability of a specific class instance, or a label value. Let $P(y=0)$ and $P(y=1)$ represent the probability of cat and dog images respectively. Now $P(X|y=0;w)$ means the probability distribution of image features given cats, and $P(X|y=1;w)$ means that of dogs. while $P(y=0|X=x;w)$ and $P(y=1|X=x;w)$ indicates the probabilities of cat and dog class labels when observing the input $X=x$.

神经网络的概率解释
Interpretation of Neural network from the probabilistic perspective

- 为适配上述解释对网络进行的更改：输出概率

Modification to the neural network to cater to the interpretation: output probability.



神经网络的概率解释
Interpretation of Neural network from the probabilistic perspective

- 为适配上述解释对网络进行的定位：估计条件概率

  Retargeting of the neural network to cater to the interpretation: conditional probability estimation.

$$D(y|X=x)=\frac{p(X=x|y=0)P(y=0)}{p(X=x|y=1)P(y=1)}$$

习题
Problems

1. 考虑随机掷两个骰子的情况，如不区分两个骰子，求此时的采样空间。

   Consider the case for tossing two dices, calculate the sampling space (assume there is no distinguishing between two dices)

2. 期望与方差是随机变量的两个重要数字特征，请自行查阅回答其定义与计算方法

   Expectation and variance are two important numerical characteristics of random variables, please search in literature for their definitions and how to calculate them.

习题
Problems

3. 阅读相关文献，了解什么是独立同分布（i.i.d）随机变量及其在模型简化中的作用。

   Do literature survey to understand what are independently identically distribution random variables and their roles in model simplification.

4. 工程实践中，经常需要从特定分布中采样。但有时构造特定分布并不容易，阅读相关文献，了解如何从简单分布，如均一分布构造特定分布。

   In practice, it might demand to sample from a certain specific distribution. However, to construct a specific distribution is not an easy task. By doing literature survey to find out methods that how to construct complicated distributions from simple distributions such as uniform distribution.

习题
Problems

5. 某个夜晚，一辆出租车肇事后逃逸。该城市共有两家出租车公司，一家公司的出租车均为绿色（“绿色”公司），拥有出租车数量为全市出租车总数的85%；另一家公司的出租车均为蓝色（“蓝色”公司），拥有出租车数量为全市出租车总数的15%。一名目击者称肇事出租车是“蓝色”公司的。法院对目击者的证词进行了测试，发现目击者在出事当时那种情况下正确识别两种颜色的概率是80%。那么肇事出租车是蓝色的概率是多少？

   One night, a taxi fled the scene after an accident. There are two taxi companies in the city, with one company runs all green taxies and another manages all blue ones. The green ones occupy 85% of total account whilst the blue ones constitutes 15%.One witness reported that committed taxi was blue. The court tests the witness’s testimony and find that in that situation, the probability to identify the right color is around 80%. So, in this circumstance what is the chance of the fled taxi is blue?

习题
Problems

- 解答：

  根据题目知，

  - $P(G) = 0.85$：出租车是绿色的概率；
  - $P(B) = 0.15$：出租车是蓝色的概率；
  - $P(X = g|G) = 0.80$：当出租车是绿色时，肇事环境下观察到绿色的概率；
  - $P(X = g|B) = 0.20$：当出租车是蓝色时，肇事环境下观察到绿色的概率
  - $P(X = b|B) = 0.80$：当出租车是蓝色时，肇事环境下观察到蓝色的概率
  - $P(X = b|G) = 0.20$：当出租车是绿色时，肇事环境下观察到蓝色的概率

  $P(B|X = b) =?$：求当肇事环境下观察到蓝色时，是蓝色出租车的概率

习题
Problems

- Solution：

  According to the question，

  - $P(G) = 0.85$: The probability that the taxi is green;
  - $P(B) = 0.15$: The probability that the taxi is blue;
  - $P(X = g|G) = 0.80$: The probability of spotting a green taxi in the accident scene given a green taxi;
  - $P(X = g|B) = 0.20$: The probability of spotting a green taxi in the accident scene given a blue taxi;
  - $P(X = b|B) = 0.80$: The probability of spotting a blue taxi in the accident scene given a blue taxi;
  - $P(X = b|G) = 0.20$: The probability of spotting a blue taxi in the accident scene given a blue taxi;

  $P(B|X = b) =?$： To calculate the probability of a blue taxi given the observation of a blue taxi.

习题
Problems

$$P(B|X=b) = \frac{P(B, X=b)}{P(X=b)} = \frac{P(X=b|B)P(B)}{P(X=b)}$$

$$P(X=b) = P(X=b|G)P(G) + P(X=b|B)P(B)$$

$$P(B|X=b) = \frac{P(X=b|B)P(B)}{P(X=b|G)P(G) + P(X=b|B)P(B)}$$

$$= \frac{0.8 * 0.15}{0.2 * 0.85 + 0.8 * 0.15} = 0.41$$