Derin Öğrenme
Deep Learning

Hazırlayan: M. Ali Akcayol
Gazi Üniversitesi
Bilgisayar Mühendisliği Bölümü

Bu dersin sunumları, "Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016." kitabı kullanılarak hazırlanmıştır.

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Olasılık teorisi

- Olasılık teorisi, **rastgele/belirsiz olayların analizi** ile ilgilenir.
- **Yapay zeka** (artificial intelligence - AI) uygulamalarında **olasılık teorisi iki şekilde kullanılır:**
  - Olasılık teorilerini kullanarak oluşturulan **problemlerin** AI sistemleriyle **çözümünde** olasılık kurallarından faydalanılır.
  - **AI sistemlerinin** davranışının **analizinde** olasılık ve istatistik kullanılabilir.
- **Olasılık teorisi,** belirsizliğin varlığının nedenini açıklar; **bilgi teorisi,** olasılık dağılımdaki belirsizlik miktarını ölçer.
- Olasılık teorisi ilk olarak **olayların sıklığını analiz etmek için geliştirilmiştir.**
- **Bilgisayar bilimlerinin** birçok alanında **girişler kesindir** ve **deterministiktir.**
- Makine öğrenmesi **belirsiz** ve **stokastik** büyüklüklerle uğraşır.

Olasılık teorisi

- **Belirsizliğin olası 3 kaynağı vardır:**
  - **Sistemde var olan doğal stokastik özellikler** modelleniyor olabilir.
  - **Eksik gözlemleme** yapılmış olabilir. Sistem davranışını ifade eden değişkenlerin tümü gözlemlenmemiş olabilir.
  - **Eksik modelleme** yapılmış olabilir. Sistemdeki bazı bilgiler göz ardı edilebilir, bunlar belirsizliğe yol açabilir.
- Bazen **basit ancak belirsiz bir kuralı kullanmak** daha **pratiktir** ve oluşturmak daha az maliyetlidir(*Kuşların çoğu uçar.*).
- **Deterministik ve karmaşık bir kuralı oluşturmak daha maliyetlidir** ve hata halen ortaya çıkabilir (*Uçmayı henüz öğrenmemiş genç kuşlar, uçma yeteneğini kaybetmiş hasta ve yaralı kuşlar dışındaki kuşlar uçar.*).

Olasılık teorisi

- Bayesian probability:
  Mutlak doğru veya yanlış **kesinlik içeren** Bayesian olasılığıdır {0, 1} (*Doktorun hasta için grip veya değil şeklindeki kararı*).
- Frequentist probability:
  Tekrara dayalı olasılık içeren sıklık olasılığıdır [0, 1]
  (*Belirli semptomları gösteren hastaların %40 grip olma olasılığı*).
- **Olasılık teorisi,** belirsizliğe sahip bir **önermenin doğru veya yanlışlığını** tanımlamak için **biçimsel kuralları sağlar.**
- Random variable: **Rastgele** farklı **değerler alabilen değişkendir.**
- Rastgele **değişkenler, kesikli** veya **sürekli** olabilir.

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Olasılık dağılımları

Kesikli değişkenler ve olasılık kütle fonksiyonları

- Probability distribution, bir **rastgele değişkenin** veya değişken kümesinin her durumu için **nasıl değer aldığını tanımlar.**
- **Kesikli değişken** üzerinde **olasılık dağılımı** probability mass function (PMF) kullanılarak tanımlanabilir.
- **PMF,** bir random değişkenin bir durumdan başka **bir duruma geçiş olasılığını eşleştiren fonksiyondur.**
- $\boldsymbol{P(x) = 1}$ ise $\boldsymbol{x}$'e geçiş kesindir, $\boldsymbol{P(x) = 0}$ ise $\boldsymbol{x}$'e geçiş olanaksızdır.
- PMF, birden fazla değişken üzerinde işlem yapabilir.
- Bu tür olasılık dağılımlarına joint probability distribution denir.
- $\boldsymbol{P(\mathrm{x} = x, \mathrm{y} = y)}$, $\mathrm{x} = x$ ve $\mathrm{y} = y$ 'nin **eş zamanlı olma olasılığıdır** $\boldsymbol{P(x, y)}$.

Olasılık dağılımları

- Bir rastgele x değişkeni üzerinde tanımlanan **probability mass function** $\boldsymbol{P}$**, aşağıdaki özellikleri sağlamak zorundadır:**
  - $\boldsymbol{P}$**'nin domain'i**, x değişkeninin olası **tüm durumlarının kümesi** olmak zorundadır.
  - **0'dan küçük** ve **1'den büyük** olasılığa sahip durum olamaz. $\forall x \in \mathrm{x}, \boldsymbol{0 \leq P(x) \leq 1}$.
  - Tüm durumların **olasılık toplamı 1 olmalıdır** (normalized). $\boldsymbol{\sum_{x \in \mathrm{x}} P(x) = 1}$.
- Uniform distribution: k duruma sahip bir rastgele x değişkeninde tüm durumlar eşit olasılığa sahiptir.

$$P(\mathrm{x} = x_i) = \frac{1}{k} \quad \text{tüm } i\text{'ler için}$$

$$\sum_i P(\mathrm{x} = x_i) = \sum_i \frac{1}{k} = \frac{k}{k} = 1$$

Olasılık dağılımları

Sürekli değişkenler ve olasılık yoğunluk fonksiyonları

- Sürekli değişken üzerinde olasılık dağılımı probability density function (PDF) kullanılarak tanımlanabilir.
- **Probability density function p, aşağıdaki özellikleri sağlamak zorundadır:**
  - ***p'nin domain'i***, x değişkeninin olası **tüm durumlarının kümesi** olmak zorundadır.
  - **0'dan küçük** olasılığa sahip durum olamaz.
    $\forall x \in \mathrm{x}, p(x) \geq 0.$
  - Tüm durumların **olasılık toplamı 1 olmalıdır.**
    $\int p(x)dx = 1$
- PDF, belirli bir durum için değil, **bir aralık için olasılığı verir.**
- $\int_{[a,b]} p(x)dx$, x değişkeninin $[a, b]$ aralığındaki olasılığını verir.

İçerik

Marjinal olasılık

- Marginal probability: Bir **altküme üzerindeki olasılık dağılımına marjinal olasılık dağılımı** denir.
- x ve y kesikli rastgele değişkenler için olasılık dağılımı $P(x, y)$ olsun.
- Belirli bir y aralığı için $P(x)$ aşağıdaki gibi ifade edilir.

$$\forall x \in \mathrm{x}, P(\mathrm{x} = x) = \sum_{y} P(\mathrm{x} = x, \mathrm{y} = y)$$

- Sürekli değişkenler için $p(x)$ aşağıdaki gibi ifade edilir.

$$p(x) = \int p(x, y)dy$$

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Şartlı olasılık

- Bazen **bir olay olduktan sonra**, ona bağlı olarak **başka bir olayın olma olasılığını** bilmek gerekebilir (conditional probability).
- $\mathbf{y} = y$ ve $\mathbf{x} = x$ için $P(\mathbf{y} = y \mid \mathbf{x} = x)$ şeklinde gösterilir.
- Şartlı olasılık $P(\mathbf{y} = y \mid \mathbf{x} = x)$ aşağıdaki gibi hesaplanabilir.

$$P(\mathrm{y} = y \mid \mathrm{x} = x) = \frac{P(\mathrm{y} = y, \mathrm{x} = x)}{P(\mathrm{x} = x)}$$

- Yukarıdaki şartlı olasılık $P(\mathbf{x} = x) > 0$ ise tanımlanabilir (şart oluşmadan ardıl hesaplanamaz).
- Bir olayın ardıllarının hesaplanmasına intervention query denir.

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Beklenen değer

- Çok sayıda değerden oluşan büyük bir kümede değişkenlerin ortalama değerleri önemlidir.
- **Beklenen değer,** $P(x)$ dağılımına göre $f(x)$ fonksiyonunun ortalama değerini ifade eder.

$$\mathbb{E}[\mathbf{X}] = \sum_x x\, f(x)$$

$$\mathbb{E}[\mathbf{X}] = \int x\, f(x) dx$$

Beklenen değer

Örnek:

- X, (a, b) aralığında **uniform** bir dağılıma sahipse, beklenen değeri bulalım.

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ or } x > b \end{cases}$$

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x\, f(x) dx \\ &= \int_a^b x\left(\frac{1}{b-a}\right) dx \\ &= \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b dx \\ &= \frac{a+b}{2}. \end{aligned}$$

- Uniform dağılım için beklenen değer aritmetik ortalamaya eşittir.

Beklenen değer

Örnek:

- X, $[0, 1]$ aralığında **sürekli rastgele değişken** ise beklenen değeri bulalım.

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x\, f(x) dx \\ &= \int_0^1 x(2x) dx \\ &= \int_0^1 2x^2 dx \\ &= \frac{2}{3}. \end{aligned}$$

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Varyans

- **Varyans,** bir rastgele *x* değişkeninin aldığı **değerlerin değişimini gösterir (σ^2).**
- Varyans, gerçekleşen değer ile beklenen değerin farklarının karelerinin toplamının aritmetik ortalamasıdır.

$$\sigma_X^2 = V[X] = \sum_{x \in D} (x - \mu)^2 = E[(X - \mu)^2]$$

$$\sigma_X^2 = V[X] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = E[X^2] - [EX]^2$$

- **Standart sapma,** varyansın kareköküdür (σ).

Varyans

Örnek:

- Bir zar için varyans ve standart sapmayı bulalım.
- Beklenen değer;

$$\mu = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

- $(X - \mu)^2$ için beklenen değer;

x	$m(x)$	$(x - 7/2)^2$
1	1/6	25/4
2	1/6	9/4
3	1/6	1/4
4	1/6	1/4
5	1/6	9/4
6	1/6	25/4

$$E((X - \mu)^2) = V(X) = \frac{1}{6}\left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) = \frac{35}{12}$$

- Standart sapma; $D(X) = \sqrt{35/12} \approx 1.707$

Varyans

Örnek:

- X, **sürekli rastgele değişken** için varyansı bulalım.

$$f(x) = \begin{cases} \frac{3}{x^4} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x\, f(x) dx \\ &= \int_1^{\infty} \frac{3}{x^3} dx \\ &= \left[-\frac{3}{2} x^{-2} \right]_1^{\infty} \\ &= \frac{3}{2}. \end{aligned}$$

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_1^{\infty} \frac{3}{x^2} dx \\ &= \left[-3x^{-1} \right]_1^{\infty} \\ &= 3. \end{aligned}$$

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = 3 - \frac{9}{4} = \frac{3}{4}$$

- Standart sapma; $\frac{\sqrt{3}}{2}$

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

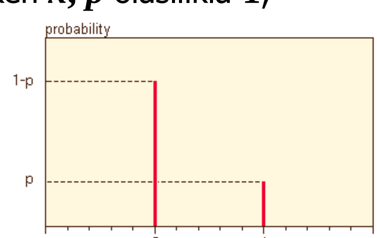
Yaygın kullanılan dağılımlar

- Olasılık dağılımları **gerçek yaşamdaki uygulamaların analizinde yaygın kullanılır.**
- Olasılık dağılımları ile olayların **sonraki gerçekleşme zamanları**, **gerçekleşme sıklıkları tahmin edilebilir.**
- İleriye dönük verilerin **ortalama değerleri, maksimum/minimum değerleri tahmin edilebilir.**
- En yaygın kullanılan dağılımlar:
  - Bernoulli dağılımı
  - Binomial dağılım
  - Poisson dağılımı
  - Uniform dağılım
  - Normal dağılım
  - Exponential dağılım

Yaygın kullanılan dağılımlar

Bernoulli dağılımı

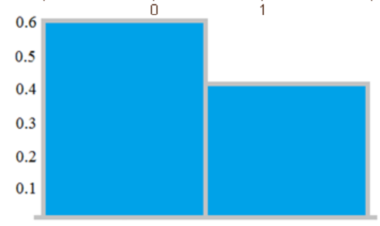
- Bernoulli dağılımı **iki olasılığa sahiptir** ve **olaylar bağımsızdır** (yazı/tura, doğru/yanlış).
- Bernoulli dağılımında, rastgele değişken $\boldsymbol{x}$, $\boldsymbol{p}$ olasılıkla $\mathbf{1}$, $(\mathbf{1}- \boldsymbol{p})$ olasılıkla $\mathbf{0}$ değerine sahiptir.



$$P(x) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \end{cases}$$

$$\mathrm{E}[X] = p \qquad \mathrm{Var}[X] = p(1-p)$$

- $\boldsymbol{p} = \mathbf{0{,}4}$ için Bernoulli dağılımı şekildeki gibidir.



Yaygın kullanılan dağılımlar

Binomial dağılım

- Aşağıdaki varsayımlar yapılmıştır:
  - ***n* adet deneme** veya test yapılmıştır.
  - Her deneme **başarılı veya başarısız** sonuçlanmıştır.
  - **Başarı olasılığı (p)** tüm denemelerde **eşit şansa sahiptir.**
  - Farklı **denemeler birbirinden bağımsızdır.**
  - n adet denemede toplam **başarılı sayısı ile ilgilenilir.**
- Yukarıdaki varsayımlar altında X **toplam başarı sayısıdır** ve binomial dağılıma sahiptir.

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

Yaygın kullanılan dağılımlar

Binomial dağılım - uygulamalar

- Aşağıdaki durumlarda binomial dağılım kullanılmaktadır:
  - Bir iş yerinde çalışan erkek/bayan sayısı
  - Başarılı satış aramalarının sayısı
  - Bir üretimde hatalı ürünlerin sayısı
  - Ağdaki bilgisayarlarda bir ayda arıza olan gün sayısı
  - Bir ağa eş zamanlı giren kullanıcı sayısı
- Ağdaki **toplam kullanıcı** sayısı $n = 35$, **eş zamanlı kullanıcı** sayısı $x = 11$, **bir kişinin** ağı kullandığı sürenin **oranı** $p = 0,1$ olmak üzere, $\mathbf{11}$ kişinin eş zamanlı ağa girme olasılığı %0,033;

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$f(11) = \frac{35!}{11!(35-11)!} 0,1^{11} (1-0,1)^{35-11} = 0,00033$$

Yaygın kullanılan dağılımlar

Binomial dağılım - örnek

- Bir sınavda 10 tane 4 seçenekli soru vardır (**n = 10** ve **p = 1/4**).
- Bir öğrenci tamamen rastgele cevaplarsa;
- Hiç doğru cevap vermeme olasılığı;

$$\begin{aligned} P(X=0) &= \frac{10!}{0!(10-0)!}(0.25)^0(1-0.25)^{10-0} \\ &= (0.75)^{10} \\ &= 0.0563 \end{aligned}$$

- İki doğru cevap verme olasılığı;

$$\begin{aligned} P(X=2) &= \frac{10!}{2!8!}(0.25)^2(1-0.25)^8 \\ &= 45 \cdot (0.25)^2 \cdot (0.75)^8 \\ &= 0.2816 \end{aligned}$$

- Testte başarısız olma (5 ve altında doğru) olasılığı;

$$\begin{aligned} P(X \le 5) &= \sum_{i=0}^{5} P(X=i) \\ &= 0.0563 + 0.1877 + 0.2816 + 0.2503 \\ &\qquad +0.1460 + 0.0584 \\ &= 0.9803 \end{aligned}$$

Yaygın kullanılan dağılımlar

Binomial dağılım

- Binomial dağılımda beklenen değer ve varyans;

$$\mu = E(X) = np$$

$$\sigma^2 = V(X) = np(1-p)$$

- 4 şıklı 10 sorudan oluşan sınav örneği için;

$$E(X) = 10 \cdot 0.25 = 2.5.$$

$$V(X) = 10 \cdot (0.25) \cdot (1-0.25) = 1.875$$

Yaygın kullanılan dağılımlar

Poisson dağılımı

- Aşağıdaki varsayımlar altında ortaya çıkar:
  - Bir zaman aralığındaki belirli sayıdaki olay ile başka zaman aralığındaki belirli sayıdaki **olay birbirinden bağımsızdır.**
  - **Bir zaman aralığındaki** belirli sayıdaki **olay dağılımı**, aynı boyuttaki **tüm zaman aralıkları için aynıdır.**
  - Zaman aralığının küçük bir parçası için **bir olayın olma olasılığı**, zaman aralığının tüm uzunluğu ile **oransaldır.**
- λ olayın olma oranı, t zaman aralığı ise, x belirlenen zaman aralığında olayın olma sayısıdır. $\mu = \lambda.t$, t aralığında ortalama olay sayısıdır. $$P(X = x) = e^{-\mu}\frac{\mu^x}{x!} \qquad x = 0, 1, 2, \ldots.$$
- Beklenen değer ile varyans;

$$E(X) = \mu \qquad V(X) = \mu$$

Yaygın kullanılan dağılımlar

Poisson dağılımı - uygulamalar

- Aşağıdaki durumlarda Poisson dağılımı kullanılabilir:
  - Bir bankaya saatlik gelen müşteri sayısı
  - Bir otoyolda günlük kaza sayısı
  - Belirli bir Web sunucuya saatlik erişim sayısı
  - Ankara’da günlük acil çağrı sayısı
  - Bir kitaptaki yazım hatası sayısı
  - Büyük bir şirkette aylık devamsızlık yapan çalışan sayısı
  - Belirli bir ürün için aylık talep sayısı

Yaygın kullanılan dağılımlar

Poisson dağılımı - örnek

- Bir kitapta **her 100 sayfada ortalama 1,5 yazım hatası** vardır **(μ = 1,5)**.
- Rastgele **100 sayfa seçildiğinde hata olmama olasılığı;**

$$P(X=0) = e^{-\mu}\frac{\mu^x}{x!} = e^{-1.5}\frac{1.5^0}{0!} = 0.2231$$

- Rastgele **400 sayfa seçildiğinde hata olmama olasılığı;**

$$\begin{aligned} P(X=0) &= e^{-1.5\cdot 4}\frac{(1.5\cdot 4)^0}{0!} \\ &= 0.002479 \end{aligned}$$

- Rastgele 400 sayfa seçildiğinde **5 ve altında hata olmama olasılığı;**

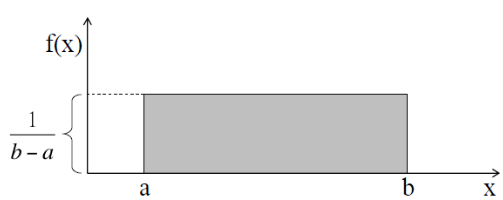
$$\begin{aligned} P(X\le 5) &= \sum_{i=0}^{5} P(X=i) \\ &= 0.0025 + 0.0149 + 0.0446 + 0.0892 \\ &\quad +0.1339 + 0.1606 \\ &= 0.4457 \end{aligned}$$

Yaygın kullanılan dağılımlar

Uniform dağılım

- ***X*** rastgele değişkeninin dağılım fonksiyonu aşağıdaki gibi ise **uniform dağılıma sahiptir:**

$$f(x) = \frac{1}{b-a} \qquad -\infty < a \le x \le b < \infty$$



- Beklenen değer ve varyans;

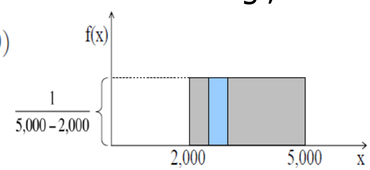
$$E(X) = \frac{a+b}{2} \qquad V(X) = \frac{(b-a)^2}{12}$$

Yaygın kullanılan dağılımlar

Uniform dağılım - örnek

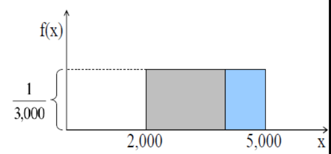
- Bir benzin istasyonu uniform dağılıma sahip olarak **günde 2000 lt - 5000 lt arasında** benzin satmaktadır.
- Bir gün için 2500 lt-3000 lt arasında benzin satma olasılığı;

$$P(2500 < X \leq 3000) = \frac{1}{5000-2000}(3000-2500) = 0.1667\,.$$



- Bir gün için en az 4000 lt benzin satma olasılığı;

$$P(X > 4000) = \frac{1}{5000-2000}(5000-4000) = 0.3333\,.$$

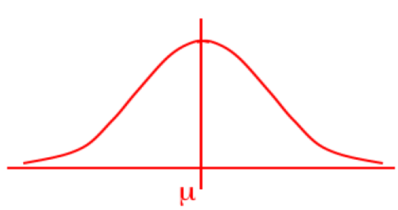


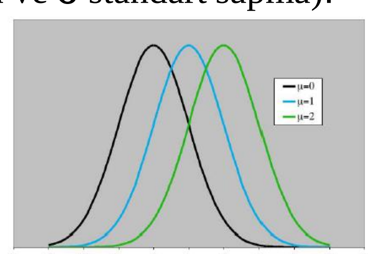
Yaygın kullanılan dağılımlar

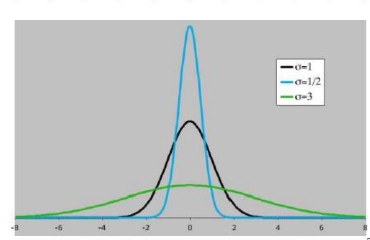
Normal dağılım

- ***X*** rastgele değişkenin dağılım fonksiyonu aşağıdaki gibi ise normal dağılıma sahiptir (**μ** ortalama ve **σ** standart sapma):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad -\infty < x < \infty$$







- Beklenen değer ve varyans;

$$E(X) = \mu \qquad V(X) = \sigma^2$$

Yaygın kullanılan dağılımlar

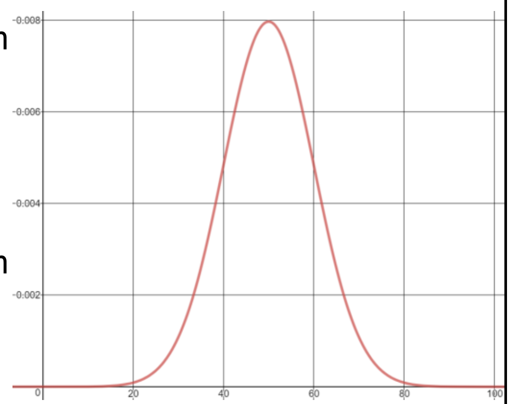
Normal dağılım - örnek

- Bir bilgisayarı toplamak **ortalama 50dk** almaktadır. **Standart sapma 10dk.**
- Yeni gelen bir bilgisayarın tam 60 dk'da toplanma olasılığı;

$$\frac{1}{10\sqrt{2\pi}} e^{-0.5\cdot\left(\frac{(60-50)}{10}\right)^2} = 0.0048$$

- Yeni gelen bir bilgisayarın tam 50 dk'da toplanma olasılığı;

$$\frac{1}{10\sqrt{2\pi}} e^{-0.5\cdot\left(\frac{(50-50)}{10}\right)^2} = 0.00798$$



Yaygın kullanılan dağılımlar

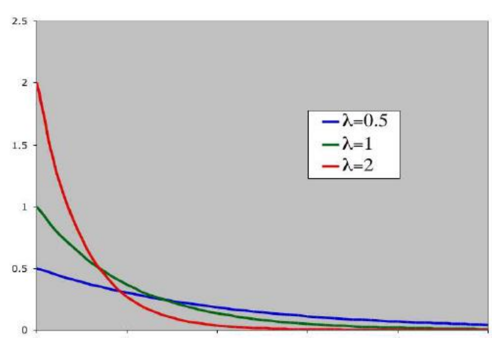
Exponential dağılım

- X rastgele değişkeninin dağılım fonksiyonu aşağıdaki gibi ise exponential dağılıma sahiptir (λ = oran):

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

- Aşağıdaki aralıklar için;

$$P\{X \leq x\} = 1 - e^{-\lambda x}$$
$$P\{X > x\} = e^{-\lambda x}$$
$$P\{x_1 < X \leq x_2\} = e^{-\lambda x_1} - e^{-\lambda x_2}$$



- Beklenen değer ve varyans;

$$E(X) = \frac{1}{\lambda} \qquad V(X) = \left(\frac{1}{\lambda}\right)^2$$

Yaygın kullanılan dağılımlar

Exponential dağılım - örnek

- Bir alkali **bataryanın ömrü X,** exponential dağılıma sahiptir ve λ = 0,05/saat.
- Bu bataryanın ortalama ömrü kaç saattir;

$$E(X) = \frac{1}{0.05} = 20$$

- Bataryanın 10-15 saat arasında bitme olasılığı nedir?

$$P(10 < X \leq 15) = e^{-0.05 \cdot 10} - e^{-0.05 \cdot 15} = 0.1341$$

- Bataryanın 20 saatten sonra bitme olasılığı nedir?

$$P(X > 20) = e^{-0.05 \cdot 20} = 0.3679$$

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Bayes kuralı

- Bayes kuralı Reverend Thomas Bayes tarafından bulunmuştur.
- Bayes kuralı, bir olayın olma olasılığına bağlı olarak, başka **bir olayın şartlı olma olasılığını hesaplamak için kullanılır.**

$$P(\mathrm{x} \mid \mathrm{y}) = \frac{P(\mathrm{x})P(\mathrm{y} \mid \mathrm{x})}{P(\mathrm{y})}$$

- Burada,
  $\boldsymbol{P}(\mathbf{x})$ x olayının olma olasılığını,
  $\boldsymbol{P}(\mathbf{y})$ y olayının olma olasılığını,
  $\boldsymbol{P}(\mathbf{y} \mid \mathbf{x})$ y olayının x olayına bağlı olma olasılığını,
  $\boldsymbol{P}(\mathbf{x} \mid \mathbf{y})$ x olayının y olayına bağlı olma olasılığını
  ifade eder.

İçerik

- Olasılık teorisi
- Olasılık dağılımları
- Marjinal olasılık
- Şartlı olasılık
- Beklenen değer
- Varyans ve standart sapma
- Yaygın kullanılan dağılımlar
- Bayes kuralı
- Bilgi teorisi

Bilgi teorisi

- Entropi, rastgele değere sahip bir değişken veya bir sistem için **belirsizlik ölçütüdür.**
- Enformasyon, rastgele bir olayın gerçekleşmesi halinde **ortaya çıkan bilgi ölçütüdür.**
- **Bir süreç için entropi**, tüm örnekler (durumlar) tarafından içerilen **enformasyonun değeridir.**
- **Eşit olasılıklı durumlara sahip sistemler yüksek belirsizliğe sahiptirler.**
- Shannon, bir sistemdeki durum değişikliğinde, **entropideki değişimin enformasyon boyutunu tanımladığını** öne sürmüştür.
- Buna göre bir sistemdeki **belirsizlik arttıkça**, bir durum gerçekleştiğinde **elde edilecek enformasyon boyutu da artacaktır.**

Bilgi teorisi

- Shannon bilgiyi bitlerle ifade ettiği için, logaritmayı 2 tabanında kullanmıştır ve enformasyon formülünü aşağıdaki gibi vermiştir.

$$I(x) = \log \frac{1}{P(x)} = -\log P(x)$$

- $P(x)$, x olayının gerçekleşme olasılığını gösterir.
- Shannon'a göre entropi, **iletilen bir mesajın taşıdığı enformasyonun değeridir.**
- Shannon entropisi H, aşağıdaki gibi ifade edilir:

$$H(X) = E(I(X)) = \sum_{1 \le i \le n} P(x_i).I(x_i)$$

$$= \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} = -\sum_{i=1}^{n} P_i \log_2 P_i$$

Bilgi teorisi

Örnek

- Bir paranın havaya atılması olayı rastsal X sürecini göstersin. Yazı ve tura gelme olasılıkları eşit olduğundan elde edilecek enformasyon,

$$I(X) = \log \frac{1}{P(X)} = \log \frac{1}{0,5} = \log 2 = 1$$

olur. **Bu olayın sonucunda 1 bitlik bilgi kazanılmıştır.**

- Entropi değeri ise 1 olarak bulunur.

$$H(X) = -\sum_{i=1}^{2} p_i \log_2 p_i$$
$$= -(0.5\log_2 0.5 + 0.5\log_2 0.5) = 1$$

Bilgi teorisi

Örnek

- Aşağıdaki 8 elemanlı S kümesi verilsin.

$S = \{evet,\ hayır,\ evet,\ hayır,\ hayır,\ hayır,\ hayır,\ hayır\}$

- ***"evet"* ve *"hayır"* için olasılık,**

$$p\,(evet) = \frac{2}{8} = 0,25 \qquad p\,(hayir) = \frac{6}{8} = 0,75$$

- **Entropi değeri,**

$$H(S) \quad = p(evet)\log_2 \frac{1}{p(evet)} + p(hayir)\log_2 \frac{1}{p(hayir)}$$
$$= 0,25.\log_2 \frac{1}{0,25} + 0,75.\log_2 \frac{1}{0,75}$$
$$= 0,81$$

Ödev

- Bayes kuralının clustering için uygulamasını içeren SCI/E dergilerinde yayınlanmış bir makale hakkında ödev hazırlayınız.