Örüntü Tanıma

Emir Öztürk
Oğuz Kırat

Prior ve Posterior Olasılık

- Prior (Önsel) Olasılık
  - Veriler gözlemlenmeden önceki bilgi değeri
  - Sistemin ilk olasılıkları
  - Rastgele ağırlık seçimi
  - $P(w)$
- Posterior (Sonrasal) Olasılık
  - Verilerin gözlemlendikten sonraki olasılık
  - Prior olasılığa bağlıdır.
  - Verilerin ne kadar olası olduğuna bağlıdır (likelihood)
  - $P(w|D)$

Prior ve Posterior Olasılık

- Prior (Önsel) Olasılık
  - İlk inanç olduğu için seçilen bir dağılım kabulü yapılabilir.
  - Örneğin model ağırlıklarının gaussian bir dağılıma sahip olduğu kabul edilebilir.
- Posterior (Sonrasal) Olasılık
  - Verilerin gözlemlendikten sonraki olasılık
  - Model parametrelerinin dağılımının yeni değer olarak belirlenmesi

Prior ve Posterior Olasılık

- Örneğin ev fiyatları tahmin eden bir model olsun
- Evin boyutu ile fiyat arası normal bir dağılım olduğu kabul edilebilir (prior)
- İlk ağırlık değerleri buna göre verilebilir
- Ayarlanan son model parametreleri (posterior) kabul edilebilir

Likelihood (Olabilirlik)

- Gözlem verisine bağlı olarak mevcut parametrelerin olasılığını ifade eder
- Verilen parametre w altında D gözlemleme olasılığı
  - $L(w) = P(D|w)$
- Şeklinde tanımlanır.
- Burada $L(w)$ fonksiyonu w değerinin alacağı her değer için ne kadar «olabilir» olduğunu ölçer.

Likelihood (Olabilirlik)

- Bir sınıflandırma ya da regresyon probleminde model çıktısı bir likelihood'dur.
- Bu olabilirlik değerinin model parametreleri ile birleştirilmesi ile posterior elde edilir
- Bunun için genellikle kullanılan yöntem negatif logaritmik hale getirilir.
- Bayes için:
- $-logP(w|D) = -logP(D|w) - logP(w) + c$

Maximum likelihood

- Maximum Likelihood yada MLE (Estimation) verilere en iyi uyan parametrenin bulunması işlemidir.
- Likelihood fonksiyonunu maksimize eden θ değerinin aranmasıdır.
  - $\mathrm{w} = \underset{\mathrm{w}}{\operatorname{argmax}}\, L(\mathrm{w})$
- Yalnızca $-logP(D|w)$ kısmı alınır.
- Model D verisi için w ağırlıklarını kullanır.
- Daha sonra $-logP(D|w)$ hesaplanır.
- Backpropagation işleminde kaybın gradyanı hesaplanır ve bulunan fark ağırlıklara yansıtılır.

Dağılımlar

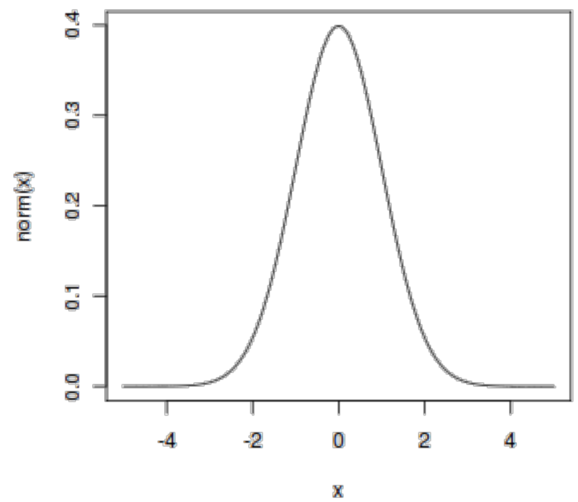
- Farklı veriler için farklı dağılımlar bulunmaktadır.
  - Normal
  - Binom
  - Poisson
  - Üstel
- Verinin dağılımını belirlemek parametre tahmininde kullanılmaktadır.
- Yanlış dağılım seçiminde eğitim başarılı olmakla beraber yeni veride hata oranı artacaktır

Dağılımlar

- Yanlış dağılım seçimi sebebi ile ağırlığa sahip olmayan verilere ağırlık verildiği için önyargı (bias) oluşacaktır.
- Böyle bir durumda overfitting ve underfitting problemleri ortaya çıkar.
- Modelin dağılımının seçilmesi makine öğrenmesinde kayıp hesaplama kısmına etki etmektedir

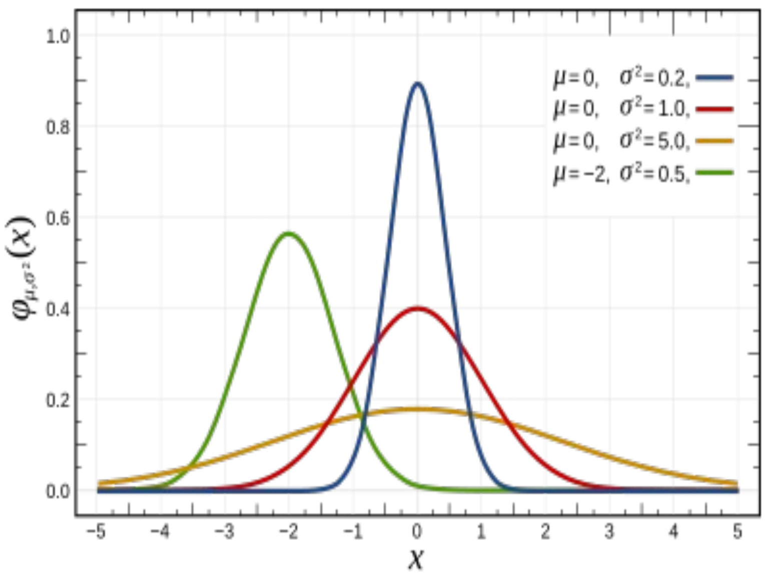
Dağılımlar

- Sürekli veri
  - Normal
  - Üstel
  - Gamma
- Ayrık veri
  - Binom
  - Poisson
  - Negatif binom



Dağılımlar

- Normal dağılımda beklenen verinin elemanlarının çoğunun ortalamaya yakın olması beklenir.
- Bire bir simetri beklenmez.
- Şiddeti farklı olabilir



Dağılımlar

- Bernoulli dağılımı
- $P(x|p) = p^x(1-p)^{1-x}$
- Tıklanmasını istediğimiz reklamlar olsun
- 0 tıklanmadı 1 tıklandı ise
- $x_1 = 1, x_2 = 0, x_3 = 1$ ise MLE?
- Normal dağılım?

Dağılımlar (Ortalamaya göre)

$$L(\mu,\sigma^2)=\prod_{i=1}^{3} f(x_i|\mu,\sigma^2)=\prod_{i=1}^{3}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right).$$

$$L(\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{3/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{3}(x_i-\mu)^2\right).$$

$$\ell(\mu,\sigma^2)=\log L(\mu,\sigma^2)=-\frac{3}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{3}(x_i-\mu)^2.$$

$$\frac{\partial\ell}{\partial\mu}=-\frac{1}{2\sigma^2}\cdot 2\sum_{i=1}^{3}(x_i-\mu)(-1)=\frac{1}{\sigma^2}\sum_{i=1}^{3}(x_i-\mu).$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{3}(x_i-\mu)=0 \quad\Longrightarrow\quad \sum_{i=1}^{3}(x_i-\mu)=0.$$

$$\sum_{i=1}^{3}x_i-3\mu=0 \quad\Longrightarrow\quad \mu=\frac{1}{3}\sum_{i=1}^{3}x_i.$$

$$\hat{\mu}=\frac{1+0+1}{3}=\frac{2}{3}\approx 0.667.$$

Dağılımlar (Varyansa göre)

$$\ell(\mu, \sigma^2) = -\frac{3}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{3}(x_i - \mu)^2.$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{3}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{3}(x_i - \mu)^2.$$

$$-\frac{3}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{3}(x_i - \mu)^2 = 0.$$

$$-3\sigma^2 + \sum_{i=1}^{3}(x_i - \mu)^2 = 0.$$

$$\sigma^2 = \frac{1}{3}\sum_{i=1}^{3}(x_i - \mu)^2.$$

$x_1 = 1$: $1 - \frac{2}{3} = \frac{1}{3}$, and $\left(\frac{1}{3}\right)^2 = \frac{1}{9}$.

$x_2 = 0$: $0 - \frac{2}{3} = -\frac{2}{3}$, and $\left(-\frac{2}{3}\right)^2 = \frac{4}{9}$.

$x_3 = 1$: $1 - \frac{2}{3} = \frac{1}{3}$, and $\left(\frac{1}{3}\right)^2 = \frac{1}{9}$.

$$\sum_{i=1}^{3}(x_i - \mu)^2 = \frac{1}{9} + \frac{4}{9} + \frac{1}{9} = \frac{6}{9} = \frac{2}{3}.$$

$$\hat{\sigma}^2 = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \approx 0.222.$$

MSE

- Mean squared error
- Hataların karelerinin toplamı
- Negatif veri için hataların doğru ölçülmesi amaçlanır
- Daha büyük hatalar daha fazla etki eder
- Uzaklık hesabı
- Görüntü verisi

MAE

- Mean absolute error
- MSE ile benzer
- Hataların mutlak değer toplamı
- Negatif veri için hataların doğru ölçülmesi amaçlanır
- Hataların etkisi eşittir
- Uzaklık hesabı
- Görüntü verisi

Cross Entropy

- İki olasılık dağılımının arasındaki farkı hesaplar
- Entropy - Redundancy
- Sınıflandırma için kayıp olarak kullanılır
- Softmax ile birlikte tahmin edilen değerlerin uygunluğunu ölçer
- $H(p, q) = -\sum_i p(i) log q(i)$