Makine Öğrenmesi

SD413

Bayes Sınıflandırıcılar

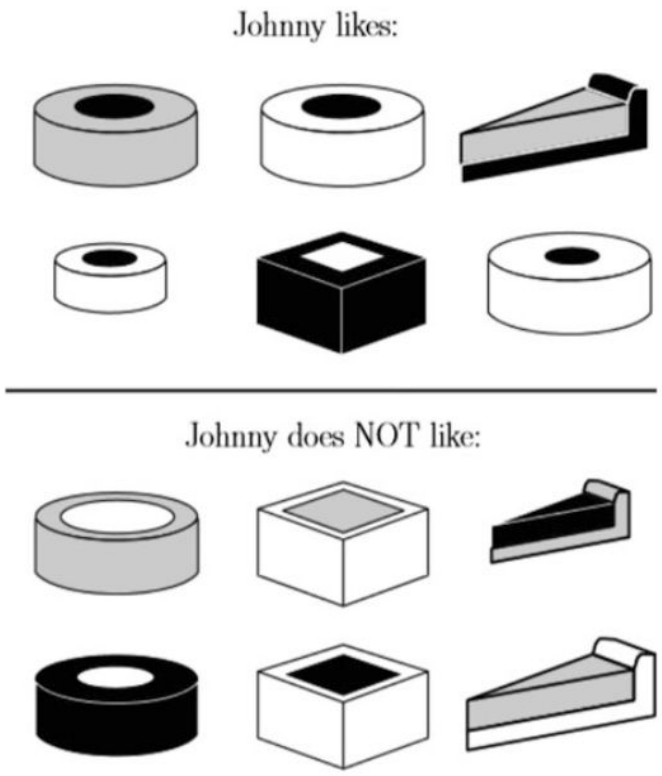
- Nitelik değerlerinin bilgisinden bir örneğin sınıfını tahmin etmeye yönelik ilk girişimler, II. Dünya Savaşı'nın çok öncesine dayanmaktadır. Tabii ki, o günlerde hiç kimse “makine öğrenmesi terimini kullanmadı, ancak amaç temelde aynıydı.

- Strateji, verilen nesne (nitelik vektörü) için, onun bireysel sınıflara ait olma olasılıklarını hesaplamak ve en yüksek olasılığa sahip sınıfı belirlemekti.

Tek Nitelikli Durum

- Gerçekçi olamayacak kadar basit bir durumla başlayalım: Her örnek tek bir nitelikle tanımlansın.
- Bir önceki haftada verilen veri tablosunu hatırlayalım.
- Eğitim seti, altısı verilen sınıfın olumlu örnekleri (N_{pos} = 6) ve altısı olumsuz (N_{neg} = 6) olmak üzere on iki turtadan (N_{all} = 12) oluşur.
- Örneklerin verilen domain'i tam olarak temsil ettiğini varsayarsak, Johnny'nin rastgele seçilmiş bir pastayı beğenme olasılığı %50 çünkü eğitim örneklerinin yarısı olumlu.

$$P(\text{pos}) = \frac{N_{pos}}{N_{all}} = \frac{6}{12} = 0.5$$



- Şimdi özniteliklerden birini seçelim, örneğin Filling-size (dolgu boyutu). Eğitim seti, üçü pozitif (N_{pos}|thick = 3) olarak etiketlenmiş, kalın dolgulu (N_{thick} = 8) sekiz örnek içermektedir.
- Dolgu boyutu=kalın olması koşuluyla bir örneğin pozitif olma koşullu olasılığının %37,5 olduğunu söylüyoruz: kalın dolgulu örnekler arasında pozitif örneklerin bağıl sıklığı şöyle hesaplanır:

$$P(\mathtt{pos}|\mathtt{thick}) = \frac{N_{pos|thick}}{N_{thick}} = \frac{3}{8} = 0.375$$

Koşullu Olasılık (Conditional Probability) ve Sınıflandırma

- Pozitif sınıfın bağıl frekansı sadece verilen öznitelik değerine sahip olan pastalar için hesaplanmıştır. Kalın dolgulu aynı sekiz pastadan beşi negatif sınıfa aittir,

  P(neg|thick) = 5/8 = 0.625.

- P(neg|thick) > P(pos|thick) olduğunu gözlemleyerek, Johnny'nin kalın dolgulu bir pastayı sevmeme olasılığının, onu sevme olasılığından daha büyük olduğu sonucuna varıyoruz.

Bayes Sınıflandırıcılar

- Bu nedenle, olasılık ilkesine dayalı bir sınıflandırıcı, herhangi bir örneği fill-size=thick durumu verili iken olumsuz bir örnek olarak etiketleyecektir.

Bayes Sınıflandırıcılar

- Koşullu olasılık, P(pos|thick), daha fazla bilgi kullandığı için önceki olasılık olan P(pos)'dan daha fazla güven uyandırır.

- Kız sayısı kadar erkek çocuğu olan bir gündüz bakımevinde, rastgele seçilen bir çocuğun, P(erkek) = 0,5 olasılıkla bir erkek olmasını bekleriz. Ancak birisinin çocuğa Johnny olarak bahsettiğini duyduğumuzda, bu beklentiyi artırırız çünkü bir kıza bu adla seslenme olasılığının daha düşük olduğunu biliriz: P(boy|Johnny) > P (boy).

Birleşik Olasılık (Joint Probability)

- Koşullu olasılık, aynı anda meydana gelen iki olayın ortak olasılığı ile karıştırılmamalıdır.

- Doğru gösterimi kullandığımızdan emin olmalıyız: birleşik olasılıkta terimler virgülle ayrılır, P(pos, thick); koşullu olasılıkla, dikey bir çubukla, P(pos|thick).

- P(pos, thick) örneğin pozitif ve dolgusunun kalın olma olasılığını ifade ederken, P(pos|thick) dolgu boyutu=kalın olanlardan birinin pozitif olma olasılığıdır.

Bayes Sınıflandırıcılar

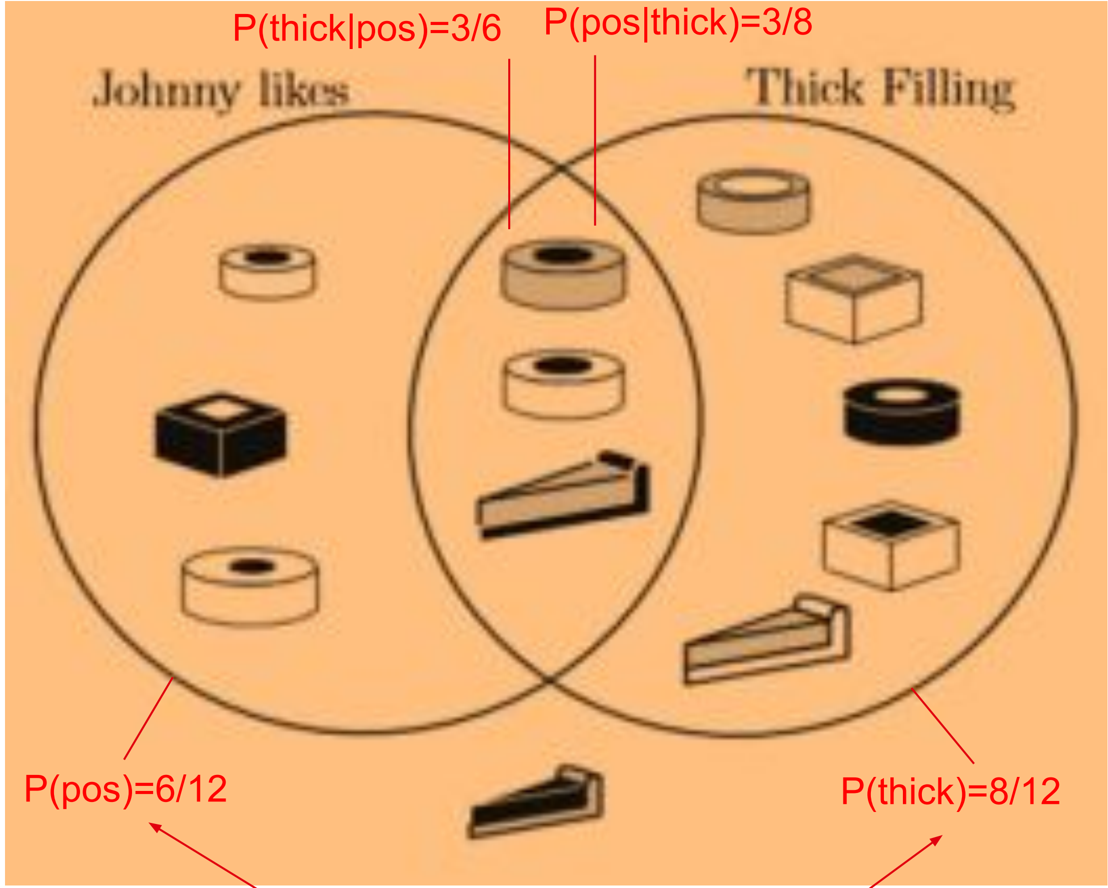
- Sonraki slaytta verilen şekil yeni tanıtılan terimleri göstermektedir.

Bayes Sınıflandırıcılar

Koşullu Olasılıklar

Birleşik Olasılık

P(pos, thick)=3/12



Önsel (Prior) Olasılıklar

Koşullu Olasılıktan Birleşik Olasılık Elde Etmek

$$P(\mathtt{pos}, \mathtt{thick}) = P(\mathtt{pos}|\mathtt{thick}) \cdot P(\mathtt{thick}) = \frac{3}{8} \cdot \frac{8}{12} = \frac{3}{12}$$

$$P(\mathtt{thick}, \mathtt{pos}) = P(\mathtt{thick}|\mathtt{pos}) \cdot P(\mathtt{pos}) = \frac{3}{6} \cdot \frac{6}{12} = \frac{3}{12}$$

Bayes Sınıflandırıcılar

- Birleşik olasılık, ilgili koşullu olasılık değerini asla aşamaz.

  P (pos, thick) ≤ P(pos|thick)

- Çünkü koşullu olasılık öncül olasılık, P (thick) veya P (pos) ile çarpılır ve bu da asla 1'den büyük olamaz.

- Ayrıca, P (thick, pos) = P (pos, thick) olduğuna dikkat edin, çünkü her ikisi de aynı örnekte oluşan kalın dolgu ve pozitif sınıfın olasılığını temsil eder. Önceki iki formülün sol tarafları eşit olduğundan, sağ tarafların da eşit olması gerekir:

$$P(\text{pos}|\text{thick}) \cdot P(\text{thick}) = P(\text{thick}|\text{pos}) \cdot P(\text{pos})$$

- Her iki tarafı da P(thick)'e bölerek, temel **Bayes formülünü** elde ederiz:

$$P(\text{pos}|\text{thick}) = \frac{P(\text{thick}|\text{pos}) \cdot P(\text{pos})}{P(\text{thick})}$$

- Tam tersi durum için de benzer bir formül elde edilebilir:

$$P(\text{neg}|\text{thick}) = \frac{P(\text{thick}|\text{neg}) \cdot P(\text{neg})}{P(\text{thick})}$$

- Bu iki formülden elde edilen değerleri karşılaştırarak, iki sınıftan hangisinin (pos of neg) doğru olma olasılığının daha yüksek olduğuna karar veririz.

- Pratik hesaplamalar bundan daha da basittir. Payda, P(thick)'i görmezden gelebiliriz.

Bayes Sınıflandırıcılar

- Koşullu olasılık ve birleşik olasılık arasındaki ilişkiden Bayes formülü nasıl elde edilir?
- Bayes formülünü bu kadar kullanışlı yapan nedir? Neyi hesaplamamızı sağlar?
- Birleşik olasılık $P(x, y)$, koşullu olasılık $P(x|y)$'den daha büyük bir değere sahip olabilir mi?
- Hangi koşullar altında $P(x|y) = P(x, y)$ olur?

Kesikli (Discrete) Özniteliklerin Vektörleri

- Şimdi, örneklerin $x = (x_1, x_2, ..., x_n)$ gibi öznitelik vektörleriyle tanımlandığı ve ikiden fazla sınıfın olduğu daha gerçekçi durumlarda Bayes formülünü kullanalım.

Çoklu Sınıflar

- Birçok gerçekçi uygulama, yalnızca pasta domain'indeki pos ve neg ile değil, ikiden fazla sınıfla işaretlenir.
- Eğer c_i, i. sınıfın etiketi ise ve x, sınıflandırmak istediğimiz nesneyi tanımlayan vektör ise, Bayes formülü aşağıdaki gibi yazılır:

$$P(c_i|\mathbf{x}) = \frac{P(\mathbf{x}|c_i)P(c_i)}{P(\mathbf{x})}$$

- Payda her sınıf için aynı olduğundan, payı maksimize eden $P(x|c_i)P(c_i)$ sınıfı seçiyoruz. Burada $P(c_i)$, eğitim setindeki c_i'nin bağıl frekansı ile tahmin edilir.

Bir Vektörün Olasılığı

- P $(x|c_i)$, c_i sınıfının rastgele seçilmiş bir örneğinin x vektörü tarafından tanımlanma olasılığıdır.
- Bu olasılığın değeri göreli sıklık ile tahmin edilebilir mi? Tam olarak değil. “Pasta” domain’inde, örnek uzayının boyutu 108 farklı örnekti, Eğitim seti bunlardan 12’sini içeriyordu, diğer vektörlerin büyük çoğunluğu hiç temsil edilmedi.

Bayes Sınıflandırıcılar

- Bağıl frekans, x'in pozitif olma olasılığının, pozitif eğitim örnekleri arasında x'i bulursak, $P(x|pos) = 1/6$, bulamazsak $P(x|pos) = 0$ olduğunu gösterir.

- Bu durumda Bayes formülündeki pay her zaman $P(x|c_i)P(c_i) = 0$ olacaktır, bu da en olası sınıfı seçmemizi imkansız hale getirir.

Bayes Sınıflandırıcılar

- Yalnızca bireysel özellikler dikkate alındığında durum iyileşir.

- Örneğin, shape=circle, pozitif örnekler arasında dört kez ve negatifler arasında iki kez meydana gelir, dolayısıyla karşılık gelen olasılıklar P (shape=circle|pos) = 4/6 ve P (shape=circle|neg) = 2/6 olur.

- Bir öznitelik yalnızca iki veya üç değer alabiliyorsa, bu değerlerin her birinin eğitim kümesinde birden fazla kez temsil edilme şansının yüksek olduğunu ve dolayısıyla olasılık tahminleri için daha iyi temeller sağladığını görüyoruz.

Naive Bayes Yaklaşımı

- Bireysel nitelik değerlerinin olasılıklarını alan ve bunları, bu değerlerin bir vektörünün belirli bir sınıfta, $P(x|c_i)$ bulunma olasılığını tahmin etmek için kullanan bir formüle ihtiyacımız var.

- Nitelikler birbirinden bağımsız olduğu sürece, bu basittir. Eğer $P(x_i|c_j)$, c_j sınıfındaki bir örneğin i. özniteliğinin değerinin x_i olma olasılığı ise, o zaman, c_j'nin rastgele bir temsilcisinin $x = (x_1, x_2, \ldots, x_n)$ ile tanımlanma olasılığı, $P(x|c_j)$, aşağıdaki gibi hesaplanır:

$$P(\mathbf{x}|c_j) = \prod_{i=1}^{n} P(x_i|c_j)$$

Naive Bayes Yaklaşımı

- Niteliklerin birbirinden bağımsız olduğu varsayımı nadiren geçerlidir.

- Gerçekten de, çeşitli değişkenlerin karşılıklı ilişkisinden kaçınılabilir mi?

- Bir nesnenin ağırlığı boyutuyla birlikte büyür, sağlık hizmetlerinin kalitesi yaşam standartlarına kadar takip edilebilir, bir bitkinin rengi belirli kimyasal özelliklerden elde edilebilir.

- Kısacası, iki özelliğin hiçbir şekilde birbirine bağımlı olmadığı alanları bulmak zordur.

Öznitelikler Arasında Karşılıklı Bağımsızlık Varsayılamazsa Ne Olur?

- Pragmatik bir yaklaşım, veri ön işleme yoluyla öznitelik bağımlılığını azaltmaya çalışmaktır.

- Değerleri başkalarından elde edilebilen gereksiz nitelikler ortadan kaldırılabilir. Örneğin, öznitelikler kümesi yaş, doğum tarihi ve şu anki tarihi içeriyorsa, bunlardan yalnızca birini kullanırsak Naive Bayes'in başarılı olma olasılığı yükselir.

- Ayrıca, iki veya üç niteliği, onları birleştiren yapay bir nitelikle değiştirebiliriz. Bu nedenle, “pasta" domain’inde, bir fırıncı bize dolgu boyutunun kabuk boyutundan bağımsız olmadığını söylemiş olabilir: biri kalınsa diğerini ince yapmayı tercih ederiz ve bunun tersi de geçerlidir. Bu durumda, iki özniteliği yeni bir öznitelikle, örneğin yalnızca iki değer alan CF-size ile değiştirmekten faydalanacağız: kalın kabuk ve ince dolgu veya ince kabuk ve kalın dolgu.

Bayes Sınıflandırıcılar

- Neden niteliklerin birbirinden bağımsız olduğunu varsaymak istiyoruz?

- Bu varsayım bize ne fayda sağlar?

- Niteliklerin birbirinden bağımsız olmadığı alanlarda bile bu varsayım neden genellikle zararsızdır?

- Niteliklerin bağımsızlığının göz ardı edilemeyeceği alanlarda neler yapılabilir?

Bilgisayar Ödevi

- Tüm özniteliklerin ayrık (discrete) olduğu bir alanda (domain) sınıf olasılıklarını hesaplamak için Bayes formülünü kullanan bir bilgisayar programı yazın. Bu programı pasta alanına uygulayın.

- Ders Sonu.