

TENSORFLOW VE PYTHON İLE METİN İŞLEME

18 EKİM 2017

BEN

Neslihan Şirin Saygılı

Fellow developer @ Prisync



sirinnnes



sirin

TAM YERİ TAM ZAMANI



TENSORFLOW GENEL BİLGİ

- Tensorflow sayısal hesaplamalar için veri akış grafikleri kullanan açık kaynaklı bir kütüphanedir.
- Aslında Google Brain Team tarafından Google'da makine öğrenmesi ve derin yapay sinir ağları araştırmaları için geliştirilmiştir, sonra genel kullanıma açılmıştır.
- Oldukça genç bir platform;
İlk sürüm: Kasım 2015
Son kararlı sürüm (1.3): Ağustos 2017
- C++ ve Python ile geliştirilmiştir.

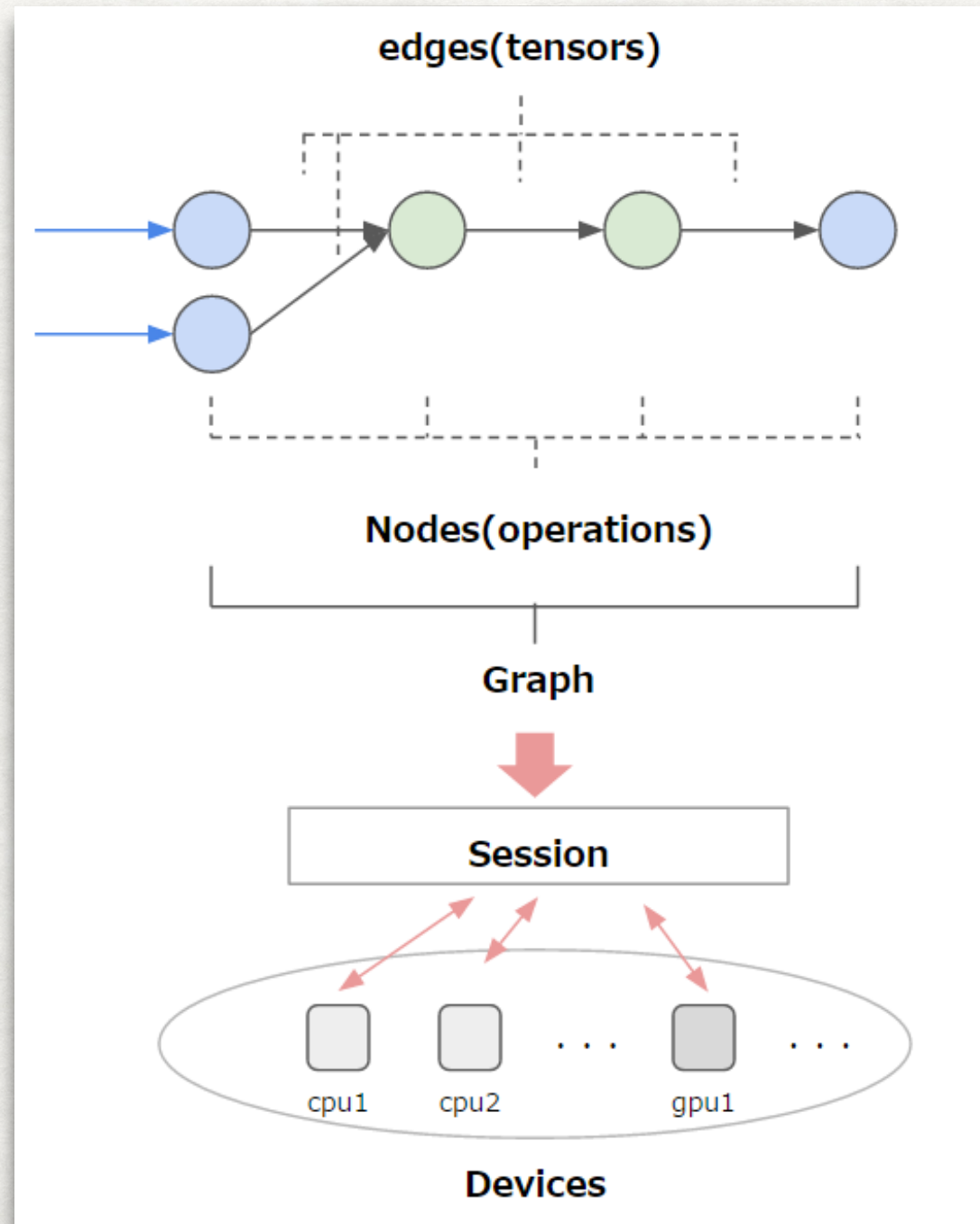
PEKİ TENSORFLOW NEDİR?

- Temel birim: **tensor**
Herhangi bir boyut sayısına göre dizilmiş bir dizi değerdir. Örneğin;
- 3 # derecesi 0 tensör
- [1., 2., 3.] # derecesi 1 olan tensör (vektör)
- [[1., 2., 3.], [4., 5., 6.]] # derecesi 2 olan tensör (matris)
- Veri akış grafiğinde düğümler matematiksel işlemleri, ayritlar tensörleri temsil eder.



PEKİ TENSORFLOW NEDİR?

- **Session:** TensorFlow çalışma zamanının denetimini ve durumunu kapsar.

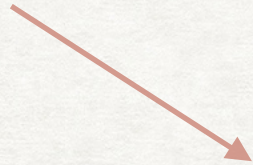


MAKİNE ÖĞRENMESİ SÜRECİ

Veri



Veriyi işlemeye
uygun hale getirme



Makine öğrenmesi
algoritması



Model



Sonuç

METİN İŞLEME ÖRNEĞİ

Problem: Belirli kategorilerde bir miktar yazı var, oluşturacağımız model ile hangi yazının hangi kategoride olduğunu tespit etmek.

Veri: Metin - 20newsgroups

Algoritma: Yapay sinir ağları - Multilayer
Perceptron

Sonuç: Metinlerin kategorileri

METİN İŞLEME ÖRNEĞİ- ADIM I

Veriyi işlemeye hazır hale getirmek

Hiç hesapta olmasa da genellikle bir hayli zaman ve uğraş isteyen işlemlerdir.

1. tüm sözcükleri al ve indexle

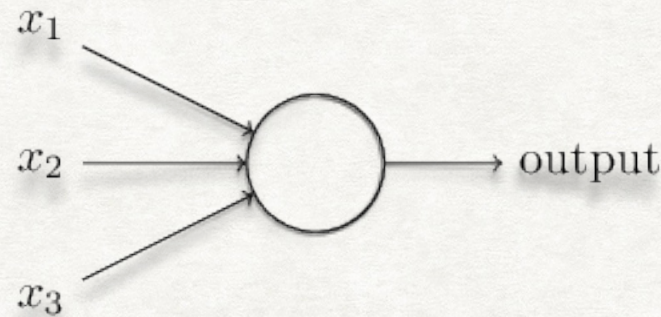
Index of the word 'memory': 4499

Index of the word 'the': 24491

2. her metin için bir matris oluşturup eğer kelime metnin içinde bulunuyorsa 1, yoksa 0 atamak.

METİN İŞLEME ÖRNEĞİ- ADIM II

Multilayer Perceptron algoritması

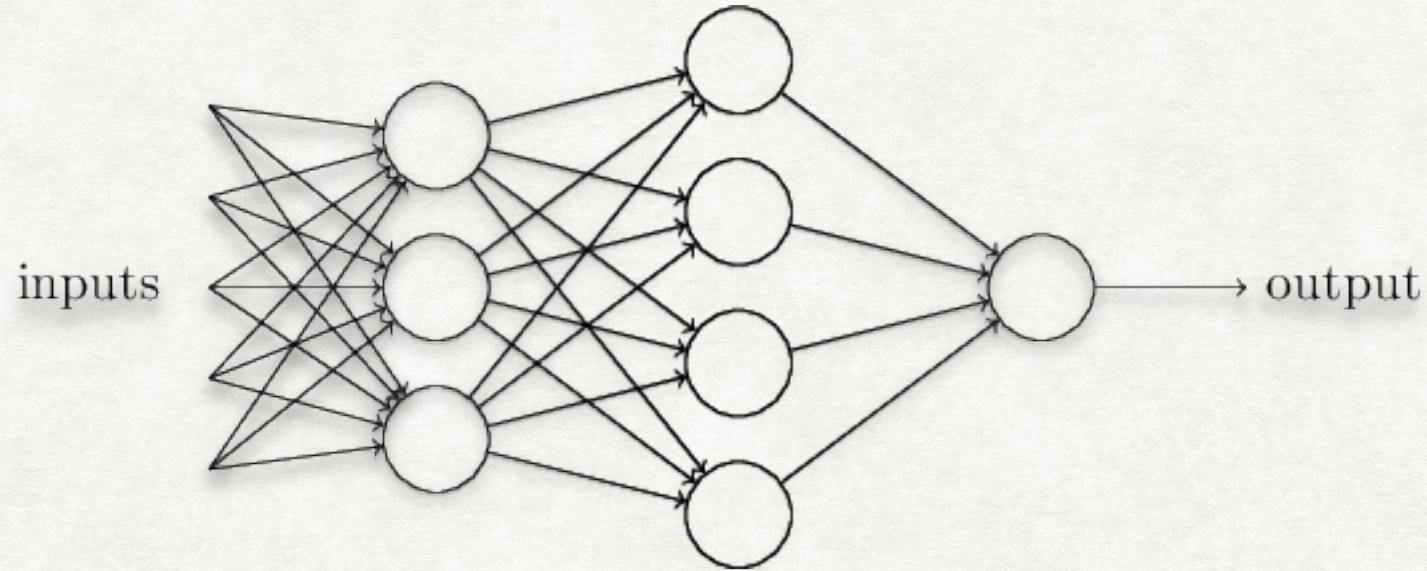


Nöron, girdiler alıp bir çıktı üreten temel birim. Her bir nöron, çıktısı 0 veya 1 olan bir aktivasyon fonksiyonuna sahiptir. Çıktı, ağırlıklı toplamın eşik değerden küçük veya büyük olup olmadığına göre belirlenir.

Aktivasyon fonksiyonu böyle olan bir nöron, bir algılayıcı (perceptron) olarak adlandırılır.

METİN İŞLEME ÖRNEĞİ- ADIM II

Multilayer Perceptron algoritması



Bu şekilde bir katmandaki nöronların çıktısı diğer katmandaki nöronlara girdi olacak şekilde meydana gelen ağa yapay sinir ağı denilebilir.

METİN İŞLEME ÖRNEĞİ- ADIM III

Model oluşturma

Her bir iterasyonda sonuç ile beklenen değeri karşılaştırıp, modelimizin optimize olması için ayarlamak gerekiyor.

Weight: verinin grafikte nasıl ilerleyeceğine dair olasılık değeri

Learning rate: weight değerlerinin ne kadar hızlı güncelleneceğini belirten parametre

Bias: Modelin veriye daha iyi uyması için kullanılan bir parametre.

METİN İŞLEME ÖRNEĞİ- ADIM IV

Sonuç

Total words: 119930

Epoch: 0001 loss= 1280.508414529

Epoch: 0002 loss= 329.705639232

Epoch: 0003 loss= 225.585577185

Epoch: 0004 loss= 65.985325532

Epoch: 0005 loss= 30.828111316

Epoch: 0006 loss= 7.401857734

Epoch: 0007 loss= 3.755576269

Epoch: 0008 loss= 0.507371463

Epoch: 0009 loss= 0.507599344

Epoch: 0010 loss= 0.286603626

Optimization Finished!

Accuracy: 0.734746

NEDEN TENSORFLOW?

- Matrix işlemlerini performans sorunu olmadan yapabiliyor.
- Doğruluk değeri yüksek sonuçlar üretiyor. (scikit-learn ile kıyasladığım birkaç denemede)
- CPU ve GPU ikisini de kullanabiliyor.
- Görselleştirme aracı TensorBoard çok gelişmiş.
- Burada yer almadı ama Deep Learning (derin öğrenme) fonksiyonları mevcut.
- Kapsamlı ve büyük projelerde cloud'da çalıştırabilmeye kolay demişler. (hacker news)

BİTİRME DEN

- Prisync, 49 ülkeden yaklaşık 400 şirkete E-ticaret alanında rekabet analizi hizmeti veriyor.
- Webden fiyat-stok bilgisi toplayıp analiz ederek anlamlı bilgi üretiyoruz.
- Prisync'te developer ekip arkadaşları arıyoruz.
- PHP - Yii Framework, Java, Python, Linux, MySQL, Bash, Redis, MongoDB, RabbitMQ, Git.
- careers@prisync.com

TEŞEKKÜRLER

- Teşekkür ederim.
- Sunumu pyistanbul.org'a ekleyeceğim.
- Soru ?