



Transfer Learning Applied to NLP and Computer Vision

15 Marzo 2019

t3chfest

Quiénes Somos

Pablo Vargas Ibarra & Manuel Lopez Sheriff

- Matemáticos UCM
- Consultores Data Scientist

K Kernel
analytics

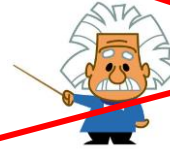


¿Qué es Machine Learning?

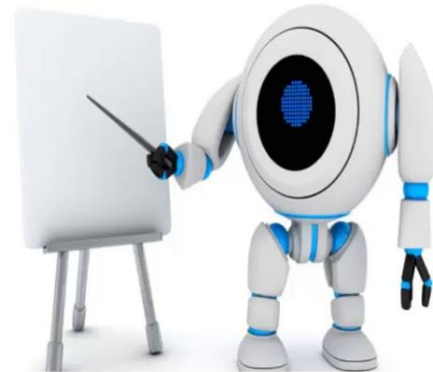
Learn From Experience



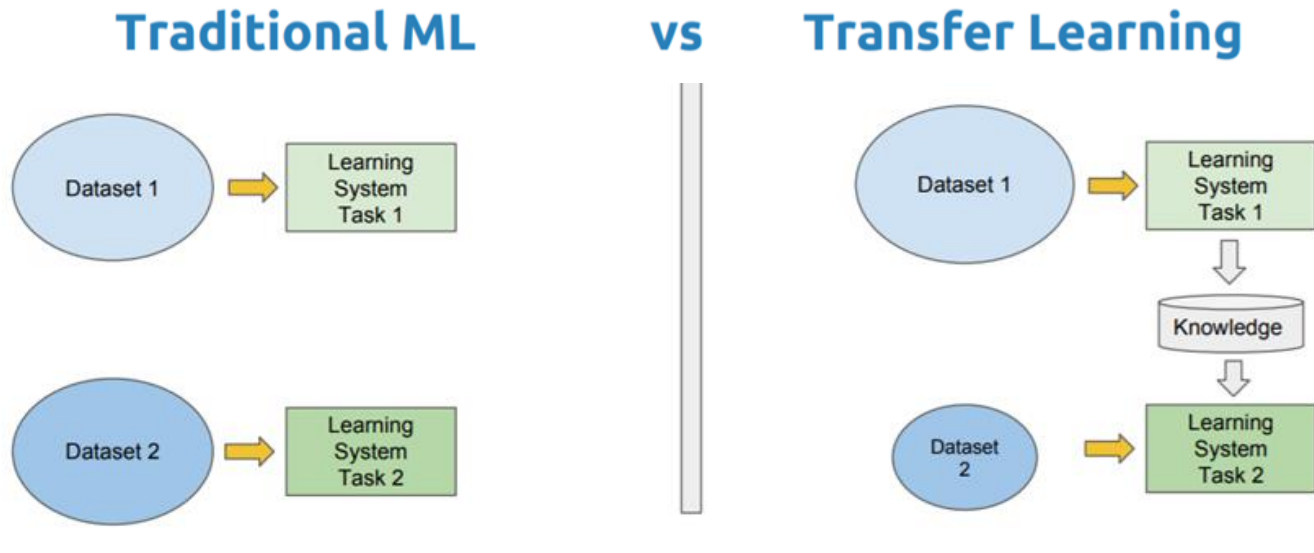
~~Reglas expertas~~



Data
Learn From ~~Experience~~



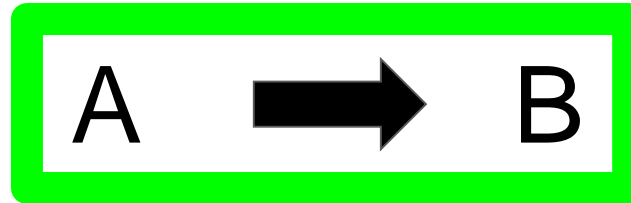
¿Qué es Transfer Learning?



After supervised learning — Transfer Learning will be the next driver of ML commercial success. Andrew Ng

Machine Learning Clásico

Aprendizaje supervisado



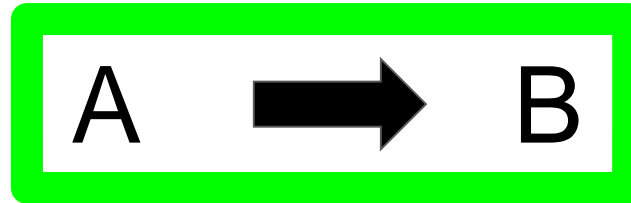
Input (A)	Output (B)	Application
email	spam? (0/1)	spam filtering
audio	text transcript	speech recognition
English	Chinese	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info	position of other cars	self-driving car

Machine Learning Clásico

Aprendizaje supervisado.

Calidad de datos = calidad modelo

Acquiring data



- Manual labeling



cat



not
cat



cat



not
cat

- From observing behaviors

user ID	time	price (\$)	purchased
4783	Jan 21 08:15.20	7.95	yes
3893	March 3 11:30.15	10.00	yes
8384	June 11 14:15.05	9.50	no
0931	Aug 2 20:30.55	12.90	yes

machine	temperature (°C)	pressure (psi)	machine fault
17987	60	7.65	N
34672	100	25.50	N
08542	140	75.50	Y
98536	165	125.00	Y

- Download from websites / partnerships

A

B

Machine Learning Clásico

Casos de uso. Problema de clasificación binaria.



- Cada ejemplo es una persona que quiere solicitar un préstamo.
- El modelo devuelve una probabilidad entre 0 y 1.
- En función del problema, se debe maximizar distintas métricas: AUC, Accuracy...



Machine Learning Clásico

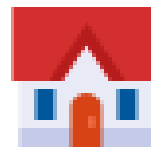
Casos de uso. Problema de regresión.



House Prices: Advanced Regression Techniques

Predict sales prices and practice feature engineering, RFs, and gradient boosting

4,106 teams · Ongoing



- La variable a predecir es continua; precio de la vivienda
- Cada ejemplo es una casa con un precio asignado
- En función de la métrica de éxito, se deben maximizar distintas métricas: MSE, RMSE, MAE, MAEN...

Machine Learning Clásico

Casos de uso. Problemas de aprendizaje NO supervisado

Ejemplos:

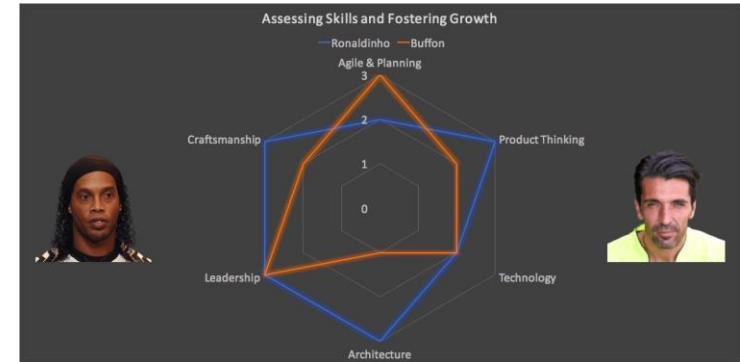
- Identificar compradores similares en el supermercado.
- Agrupar jugadores de fútbol parecidos.
- Señalar estaciones de BICIMAD semejantes.

Problemas dónde:

- Datos NO etiquetados: no hay una variable a predecir.
- Agrupamos por características similares.

Dificultades:

- Establecer distancia entre elementos $d(x, y)$.
- Elegir número de grupos (clusters).



Machine Learning

Algunos problemas difíciles

¿Cuántas personas hay en esta foto?



Traduce:

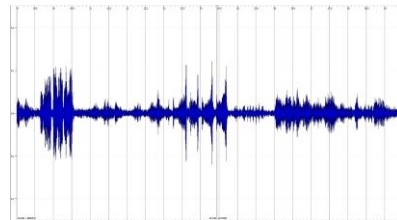
“Un banco de peces”



to english...

¿?

¿Qué género musical?



88% rock

8% punk

4% ska

¿Existe tumor en estos pulmones?



Comentarios ✓ o ✗ ?

★★★★★ Perfectas

4 de enero de 2019

Compra verificada

Perfectas y bonitas, nada más que agregar.

Útil

Comentar

Informar de un abuso



Alberto

★★★★★ Perfectas

13 de abril de 2018

Compra verificada

Son exactamente iguales a las de Parrot pero más bratas. El dr mejor con los colores.

Speech-to-text



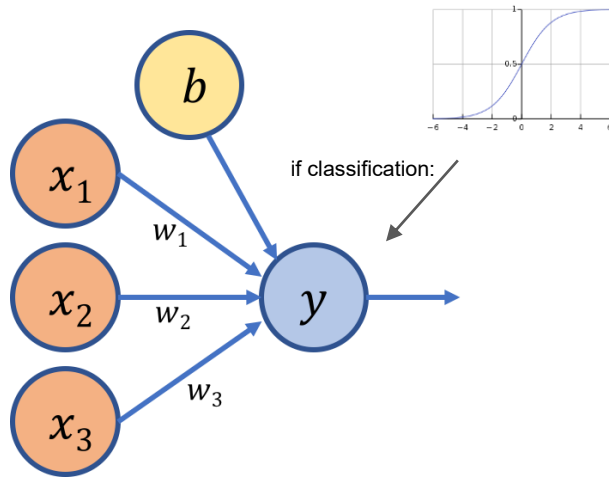
“Pablo está tranquilo”

Deep Learning

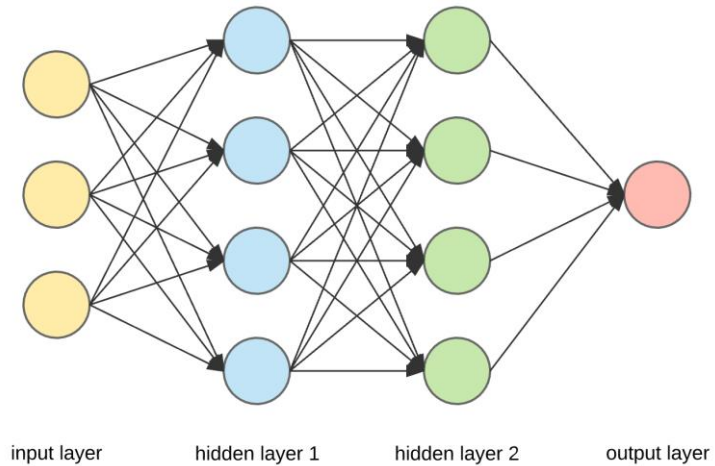
Redes neuronales

1 sola capa:

- regresión lineal
- regresión logística (clasificación)



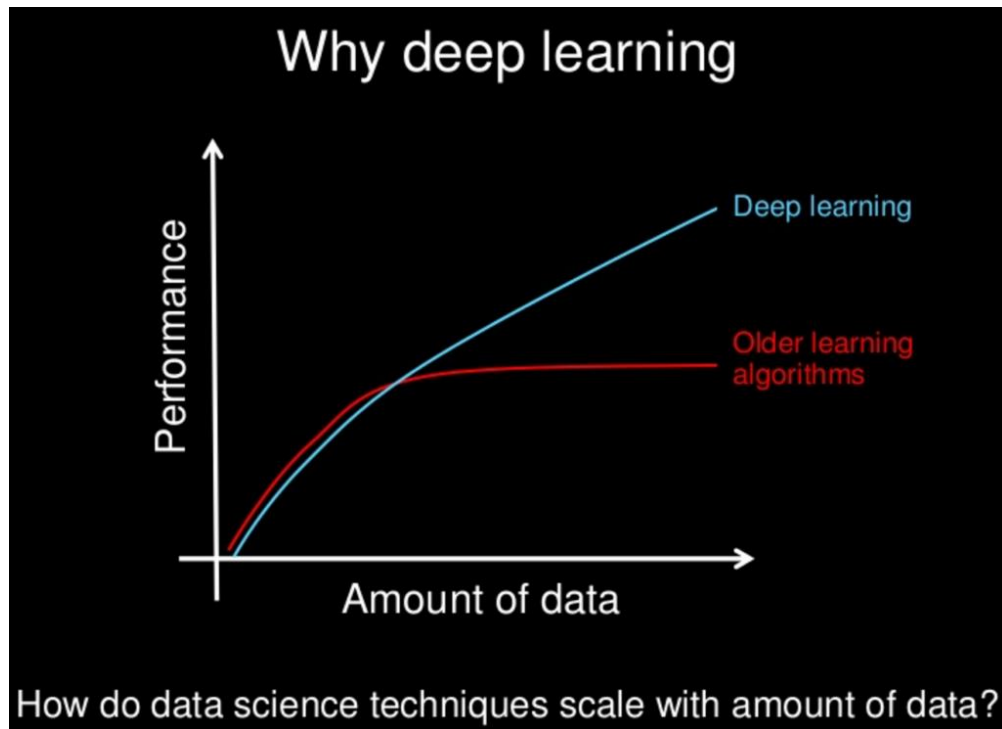
varias capas:



Deep Learning

Mejoras dentro del ML

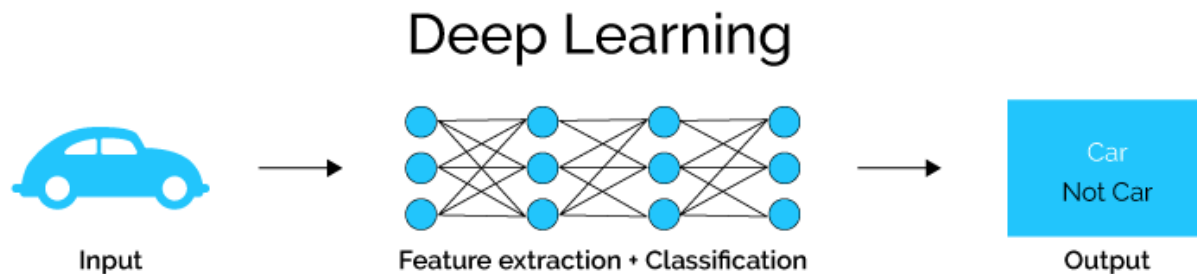
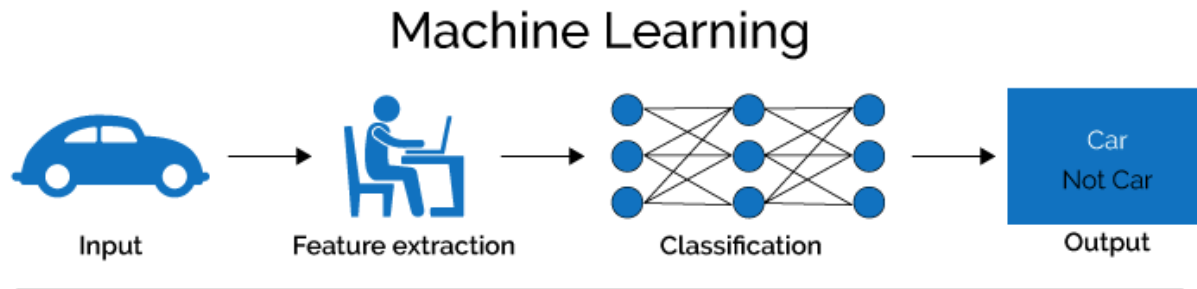
- Las NN pueden aproximar cualquier función
- Conforme aumentamos el conjunto de datos podemos conseguir una mayor precisión
- Permite un entrenamiento “online” (warm start)



Deep Learning

Diferencias con ML

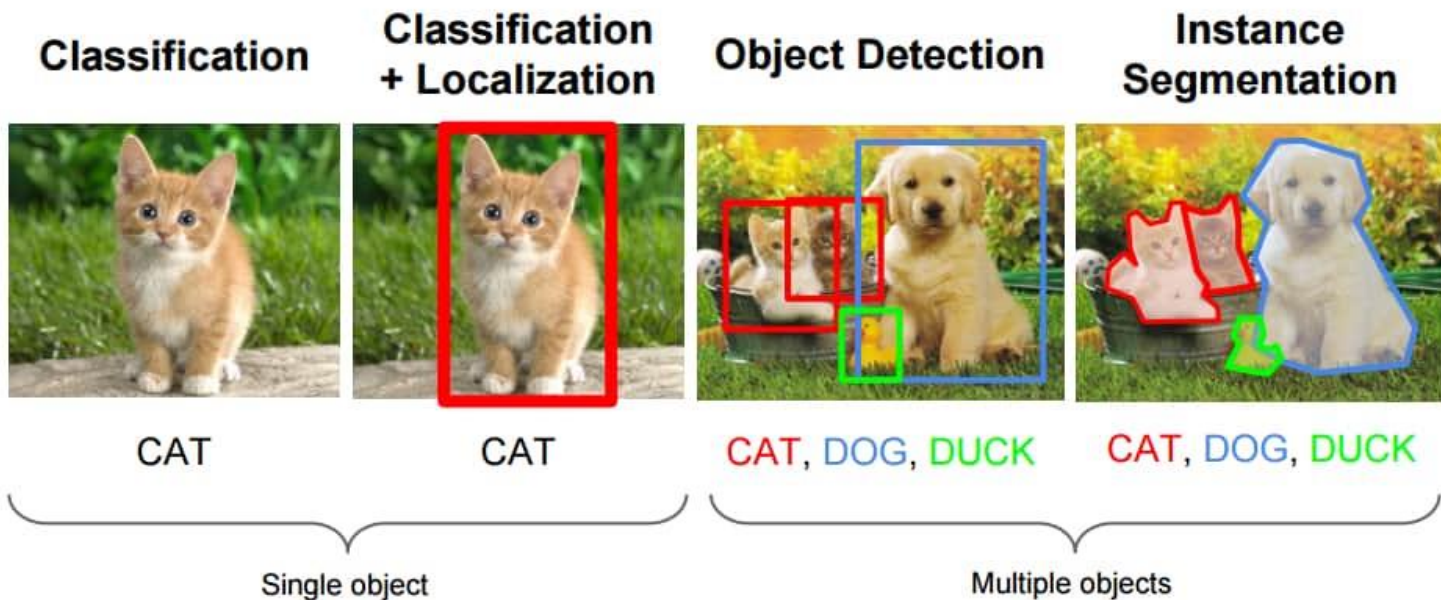
- Aprende representaciones ricas de los datos de manera automática
- Menos esfuerzo humano y más esfuerzo computacional



Computer Vision

Introducción

Computer Vision Tasks



Computer Vision

Redes convolucionales. Lectura de imágenes.

CNNs:

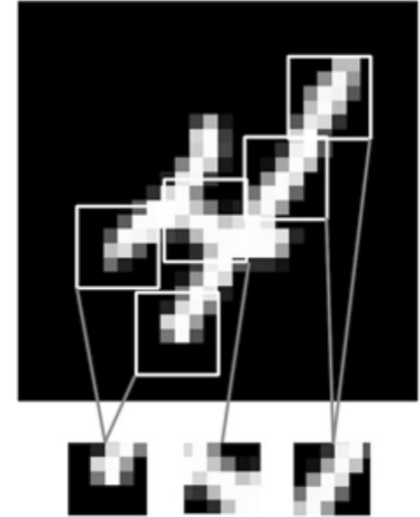
- Son invariantes traslacionales.
- Aprenden patrones locales.
- Parámetros compartidos.



Cat

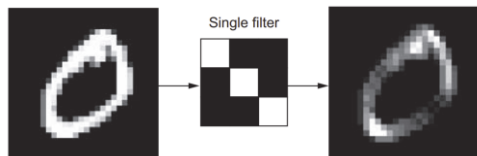


Cat

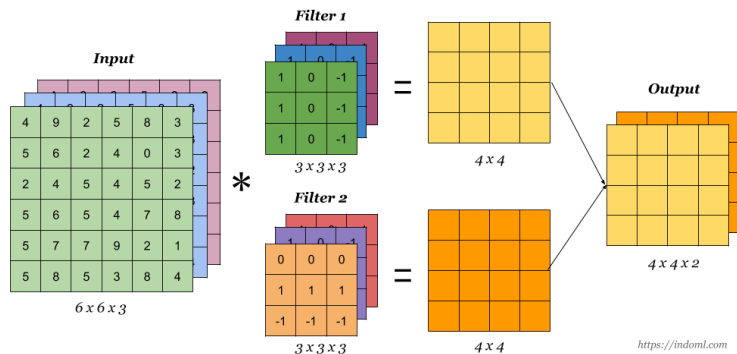


Computer Vision

Filtros



Una CNN **aprende** cuáles son los filtros ideales para el problema

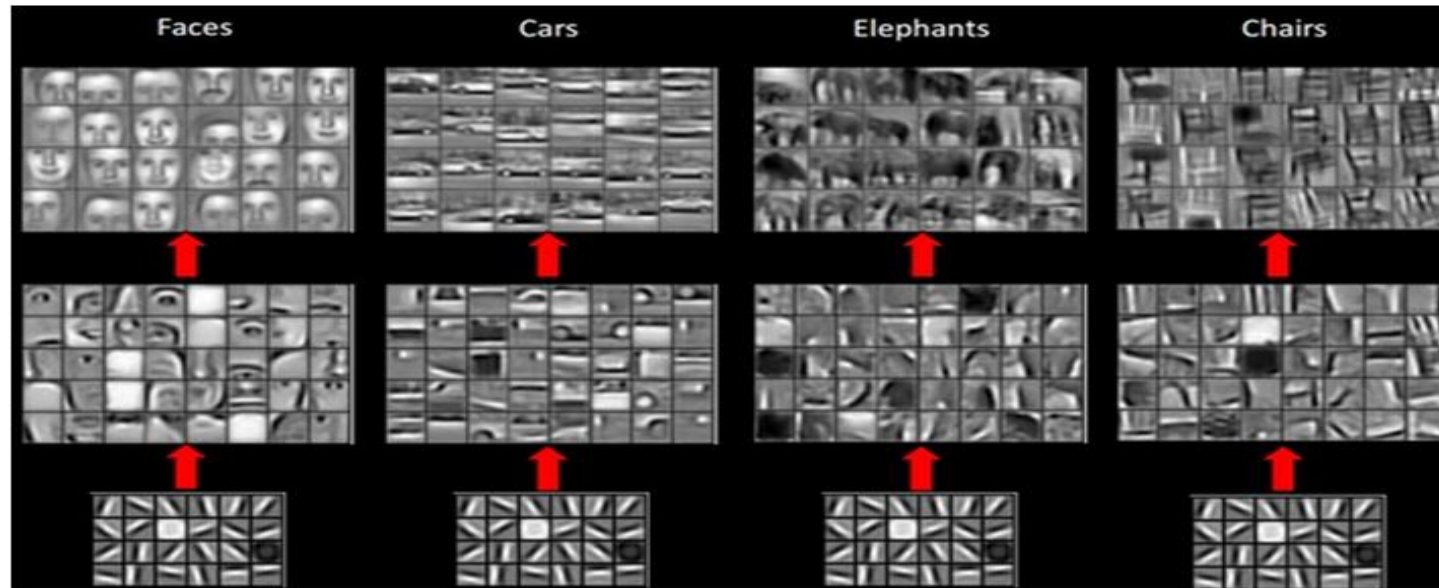


Sólo tenemos que encargarnos de prescribir la arquitectura

Layer (type)	Output Shape	Param #
reshape_45 (Reshape)	(None, 80, 3)	0
conv1d_145 (Conv1D)	(None, 71, 100)	3100
conv1d_146 (Conv1D)	(None, 62, 100)	100100
max_pooling1d_39 (MaxPooling)	(None, 20, 100)	0
conv1d_147 (Conv1D)	(None, 11, 160)	160160
conv1d_148 (Conv1D)	(None, 2, 160)	256160
global_average_pooling1d_29	(None, 160)	0
dropout_29 (Dropout)	(None, 160)	0
dense_29 (Dense)	(None, 6)	966
Total params: 520,486		
Trainable params: 520,486		

Computer Vision

Deep CNNs



Computer Vision

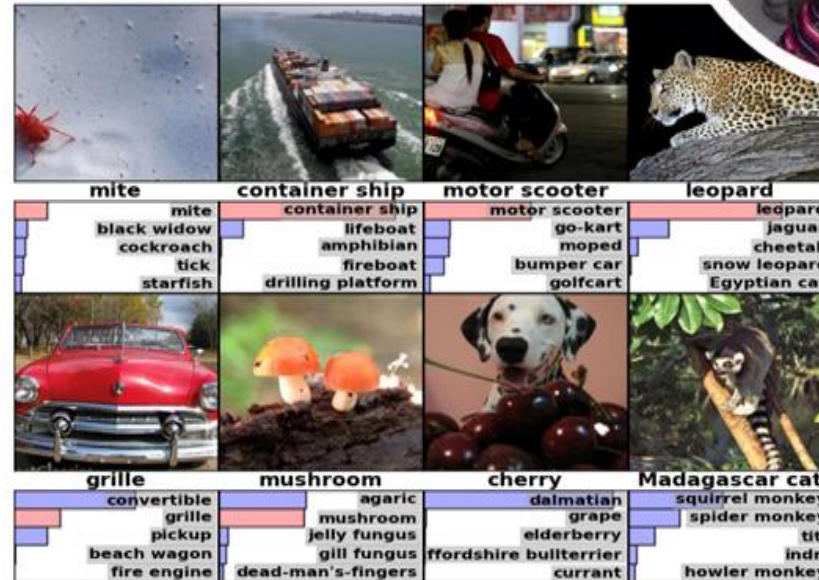
ImageNet

ImageNet Challenge



IMAGENET

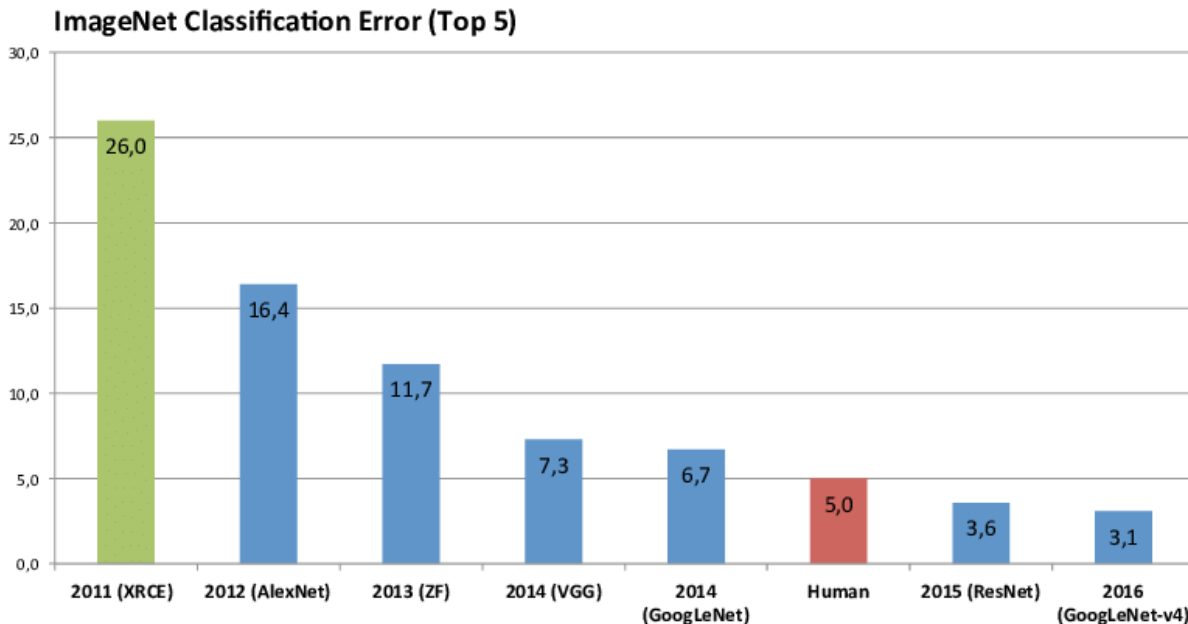
- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.



Computer Vision

Algunos resultados en el Imagenet Challenge

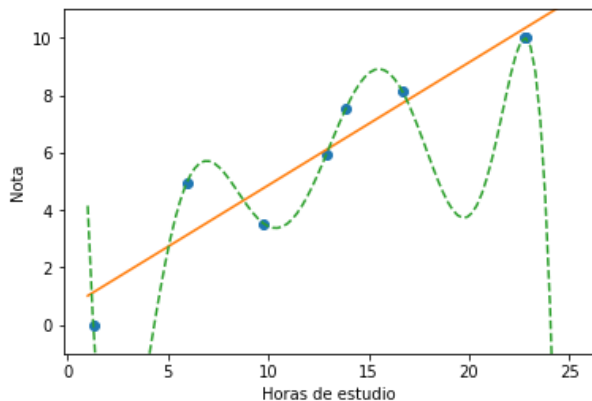
- Mejora sustancial del error cada año desde el inicio de la competición
- Precisión sobrehumana desde el año 2015
- Actualmente se ha aumentado la complejidad detectando bounding boxes



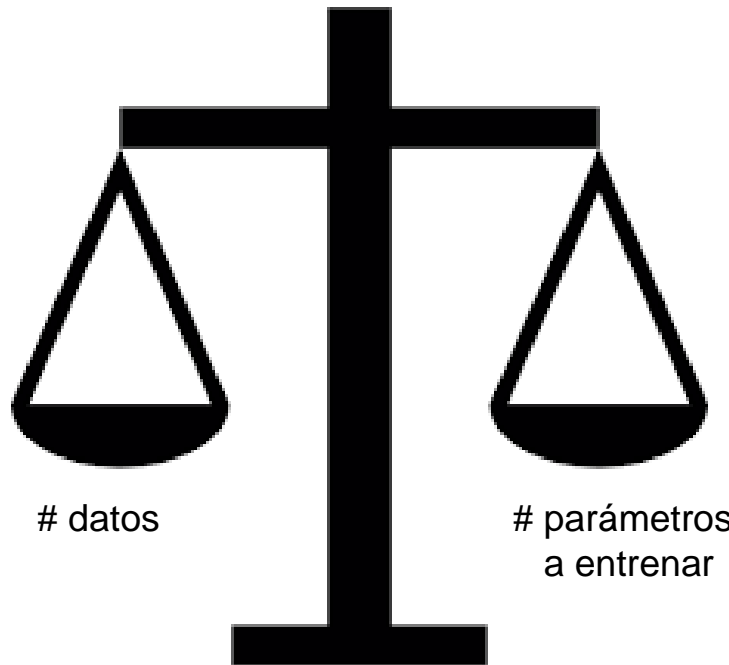
Computer Vision

Un problema: el overfitting

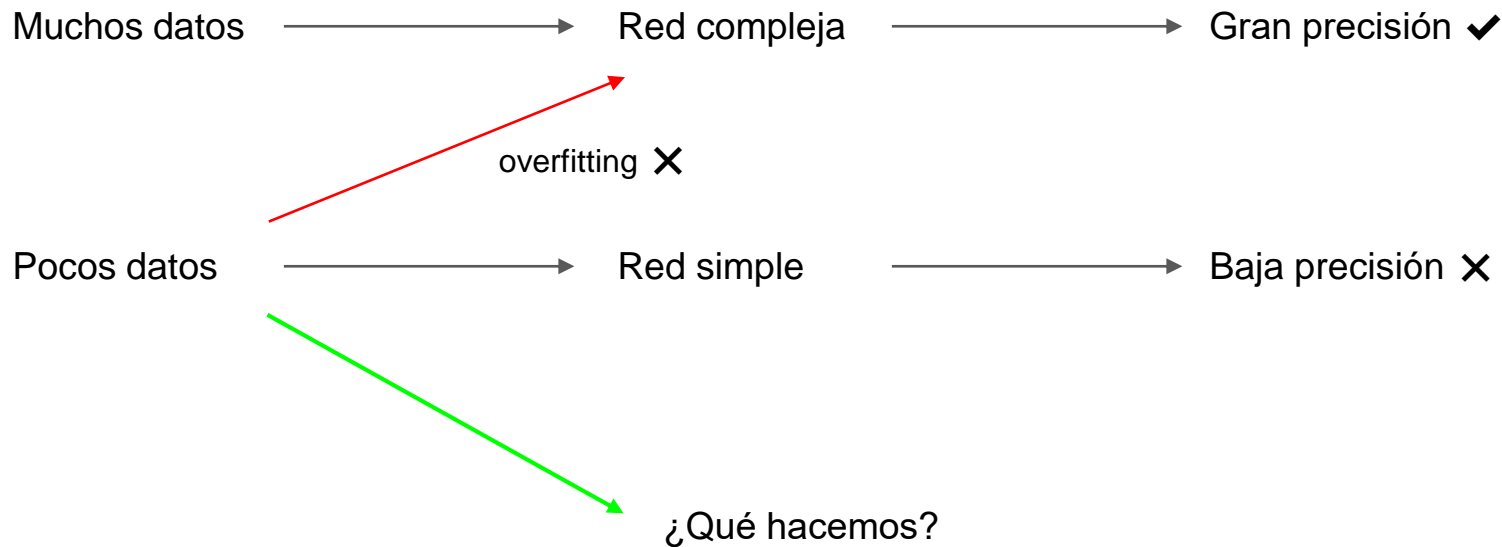
Trainable params: 520,486



Una solución: data scientist equilibrado



Computer Vision

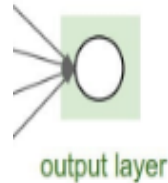


Computer Vision

Transfer Learning in Computer Vision

Nuevo conjunto de datos pequeño. Feature-Extractor.

- Modelo lineal sobre las últimas capas obtenidas de aplicar una red re-entrenada
- Previene el overfitting al tener pocos datos etiquetados para nuestro problema



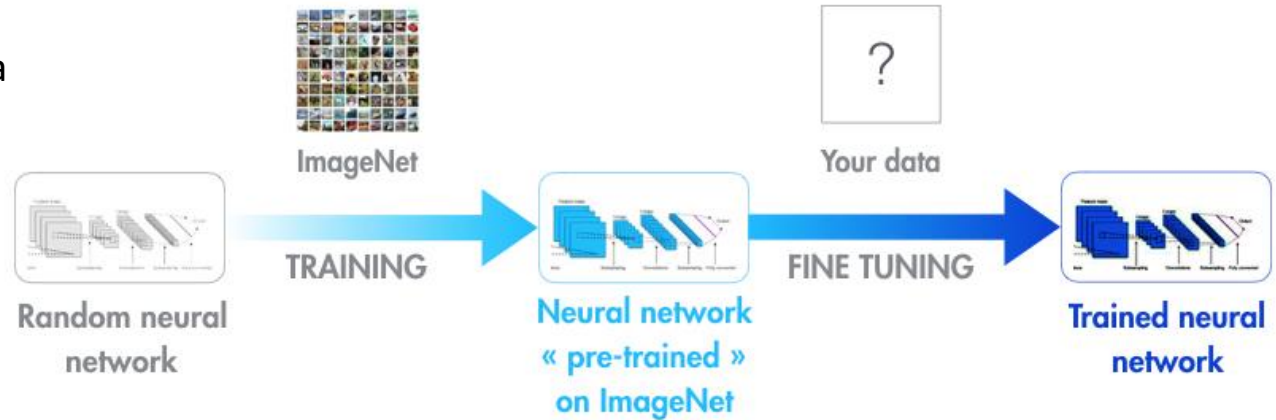
Similar dataset	Different dataset
Transfer learning: highest level features + classifier	Transfer learning: lower level features + classifier

Computer Vision

Transfer Learning in Computer Vision

Nuevo conjunto de datos grande. Fine-Tuning.

- En menos tiempo conseguimos misma o más precisión

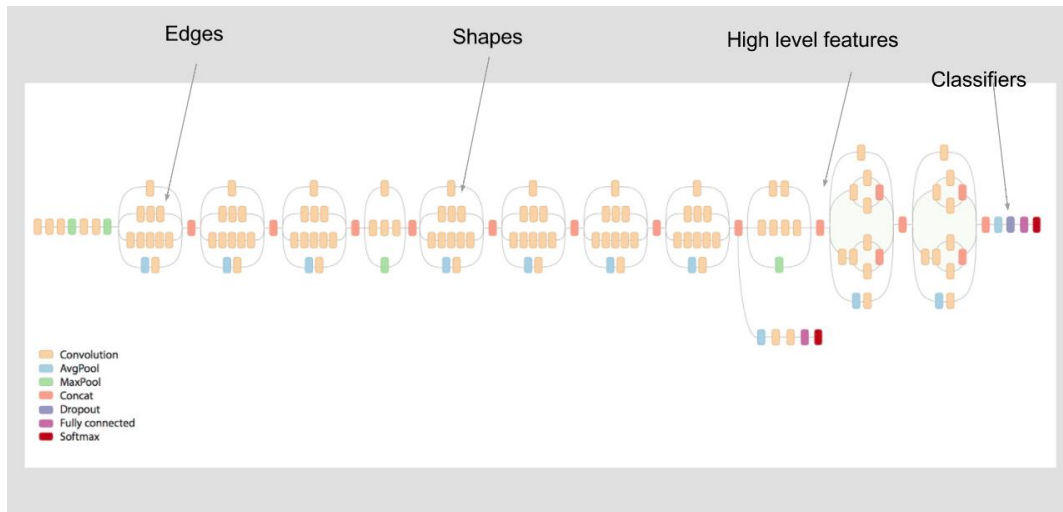


Computer Vision

Transfer Learning in Computer Vision. Summary Visual.

- Entrenamos más las capas del final dado que son específicas al problema con el que se entrenó.

What do the layers learn?



Computer Vision

Transfer Learning in Computer Vision. Example of results. Fastai



Dogs vs. Cats

Create an algorithm to distinguish dogs from cats

215 teams · 5 years ago

- Previo a la competición el estado del arte era un 80% de precisión
- Sólo 4 años después. Conseguimos una precisión de más del 98% en segundos con pocas líneas de código. (fastai transfer learning)

```
arch=resnet34
data = ImageClassifierData.from_paths(PATH, bs=16, tfms=tfms_from_model(arch, sz))
learn = ConvLearner.pretrained(arch, data, precompute=True)
learn.fit(0.01, 2)
```

39%|██████| 566/1438 [05:33<11:46, 1.23it/s]

Natural Language Processing

Introducción

Tratamos de resolver problemas relacionados con texto:

- Análisis de sentimiento: comentarios Amazon ✓ o ✗.
- Traducción: inglés → español.
- Generación de textos automática con estilo.
- Resumir un texto.

Dos desafíos:

- Representación de las palabras.
- Tratamiento de datos secuenciales.

Natural Language Processing

Primer desafío. Representación del texto.

- Siglo XX: bolsa de palabras. Frecuencia acumulada relativa.

- Lematización:

película".



"buena película" ~ "buenísima

- Sin noción de similaridad:

"buena película" != "genial película".



- Orden de palabras inutilizado:

"buen guión, mal actor" ~ "buen actor, mal guión"

$$F : \text{vocabulary} \rightarrow \mathbb{R}^N$$

- Siglo XXI, objetivo: mapear palabras en vectores. "hola" $\mapsto (0.1, 0.3, \dots, 0.7, 0.4)$

Natural Language Processing

Word Embeddings

Mapean cada palabra a un vector. Hay **peores** y **mejores** maneras de hacerlo.

	One hot encoding	Word embeddings
“gato”	(1, 0, 0, 0, ..., 0)	(0.7, 0.3, 0.1, ..., 0.9)
“perro”	(0, 1, 0, 0, ..., 0)	(0.6, 0.3, 0.7, ..., 0.4)
dimensión N	10000	300
densidad	sparse	densa
aprendido	no	sí
similarity(“gato”, “perro”)	0	0.8

Relaciones semánticas del lenguaje
embedding



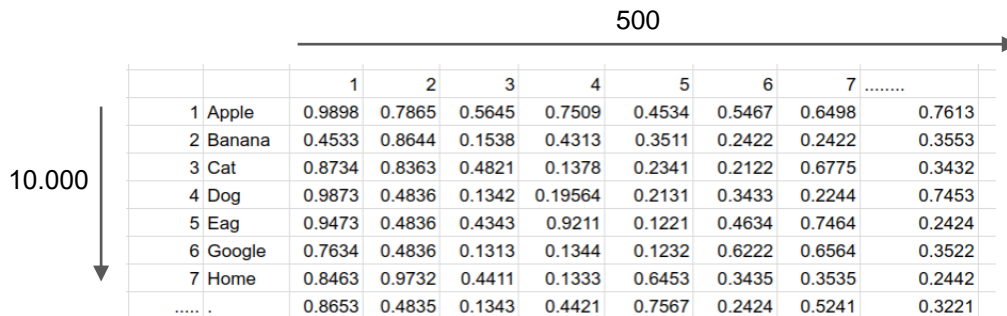
Relaciones geométricas del **word**

Natural Language Processing

Word Embeddings: gigante ejercicio.

Cada palabra de mi vocabulario simple (10k) ha de ser mapeada a un vector de (p. ej) 500 pesos

5 MILLONES de pesos



		1	2	3	4	5	6	7
1	Apple	0.9898	0.7865	0.5645	0.7509	0.4534	0.5467	0.6498	0.7613
2	Banana	0.4533	0.8644	0.1538	0.4313	0.3511	0.2422	0.2422	0.3553
3	Cat	0.8734	0.8363	0.4821	0.1378	0.2341	0.2122	0.6775	0.3432
4	Dog	0.9873	0.4836	0.1342	0.19564	0.2131	0.3433	0.2244	0.7453
5	Eag	0.9473	0.4836	0.4343	0.9211	0.1221	0.4634	0.7464	0.2424
6	Google	0.7634	0.4836	0.1313	0.1344	0.1232	0.6222	0.6564	0.3522
7	Home	0.8463	0.9732	0.4411	0.1333	0.6453	0.3435	0.3535	0.2442
.....	.	0.8653	0.4835	0.1343	0.4421	0.7567	0.2424	0.5241	0.3221

Tenemos 2 opciones:

- Crear el embedding: preciso mucho texto y computación...
- Heredar uno ya existente: glove, word2vec (2013-14), fasttext

Natural Language Processing

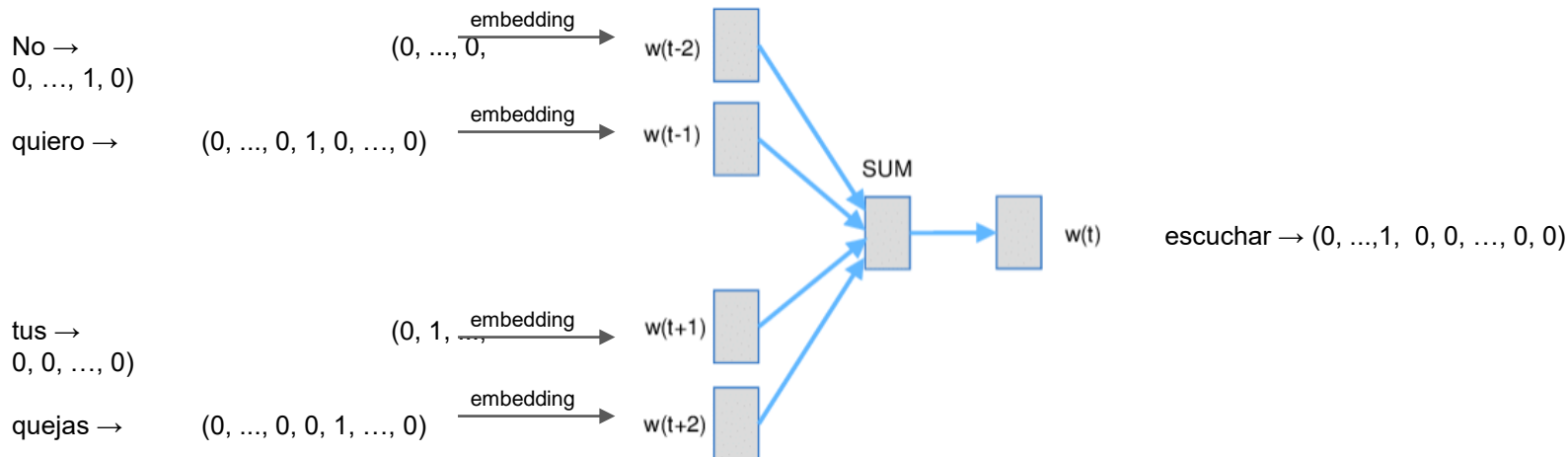
Word Embeddings: cómo se construyen

“Fui al banco a sacar _____”

- A. al perro
- B. sonrisas
- C. dinero
- D. la basura

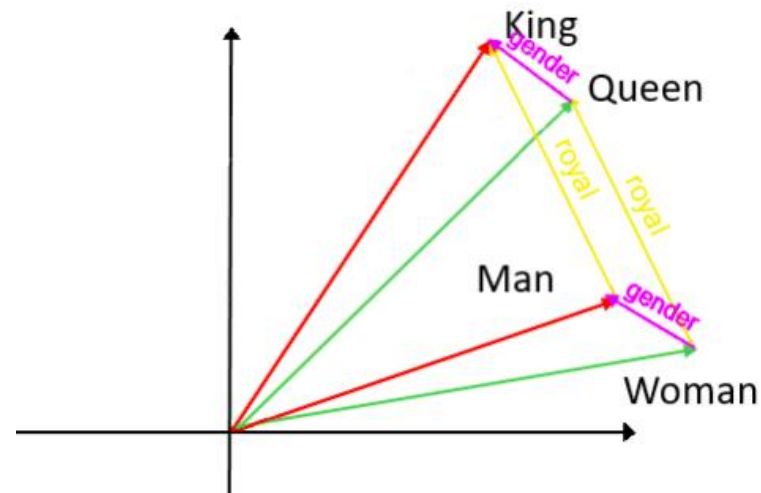
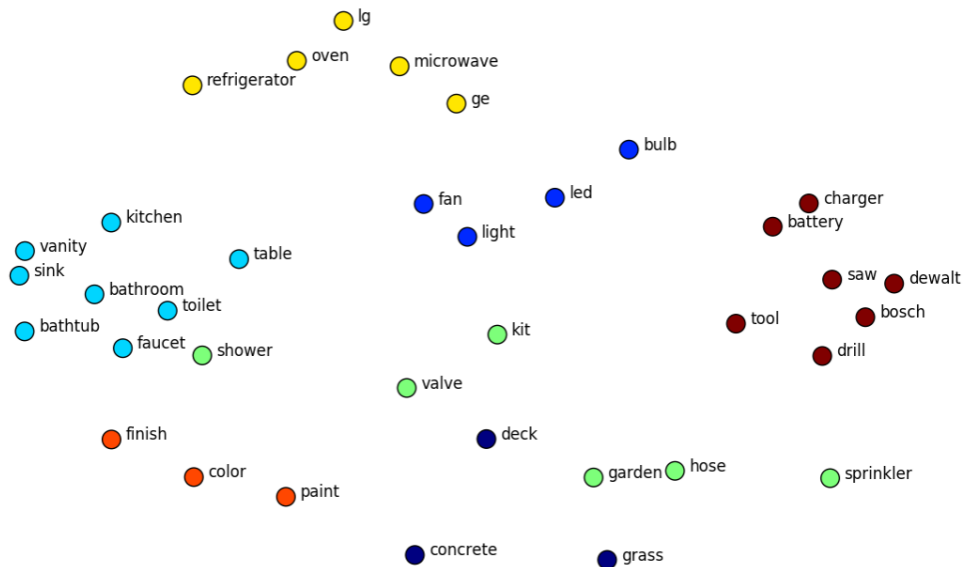
“No quiero _____ tus quejas”

- A. oír
- B. comer
- C. escuchar
- D. bailar



Natural Language Processing

Word Embeddings: cómo se construyen

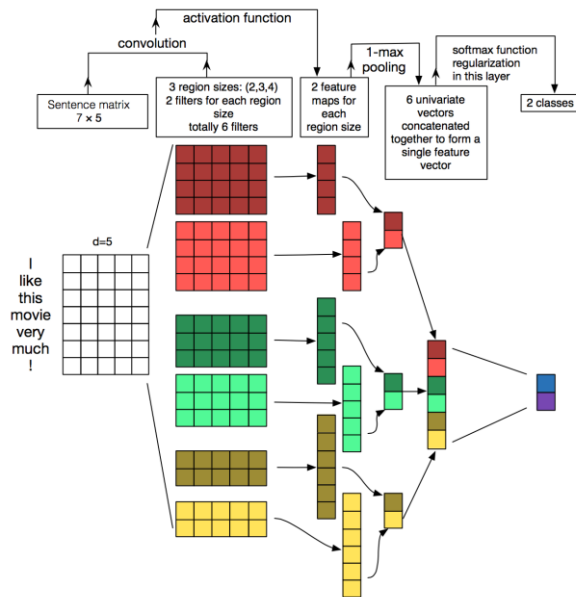


¿Sirve para mi problema?

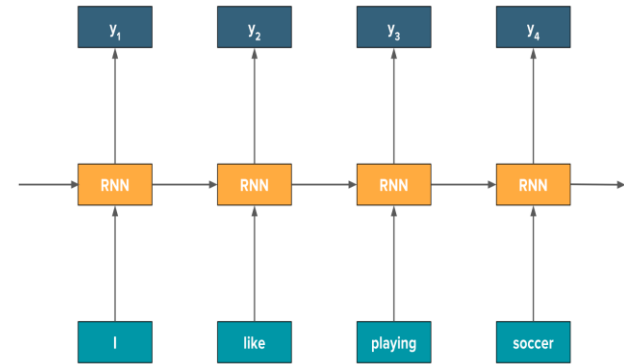
Natural Language Processing

Redes para resolver un problema de NLP

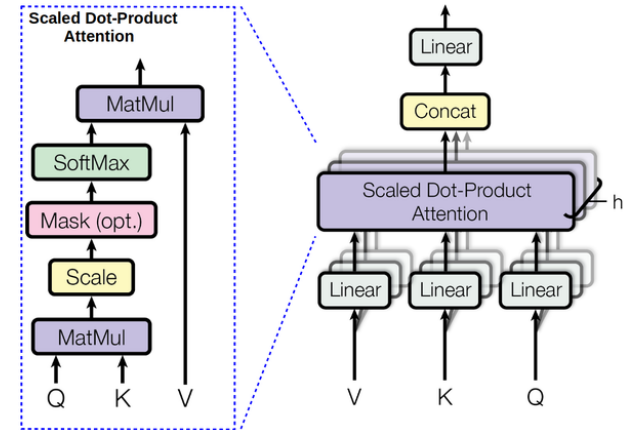
CNN



RNN



Transformer



Natural Language Processing

¿Qué nos gustaría?

- Encontrar buenas representaciones de frases / párrafos.
- Arquitecturas útiles para problemas de naturaleza distintas
- Rápidas de entrenar / aplicar fine-tuning
- Varios idiomas
- Aprendizaje de relaciones complejas (negación, subordinación ..)

Natural Language Processing

Limitación de los Word Embeddings. Palabra muñeca.



Natural Language Processing

Transfer Learning Models

BERT: Bidirectional Encoder Representations from Transformers



ELMo: Embeddings from Language Models



ULM-FiT: Universal Model Fine-Tuning for Text Classification

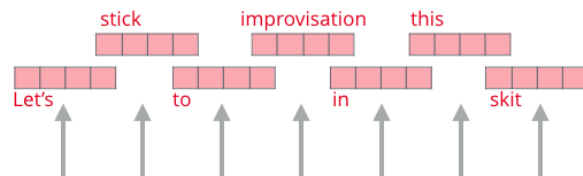


Natural Language Processing

Transfer Learning Models: ELMo (contextualized embeddings)

- Las palabras cambian de significado en función del contexto
- Deep. Representaciones más ricas de los datos.

ELMo
Embeddings



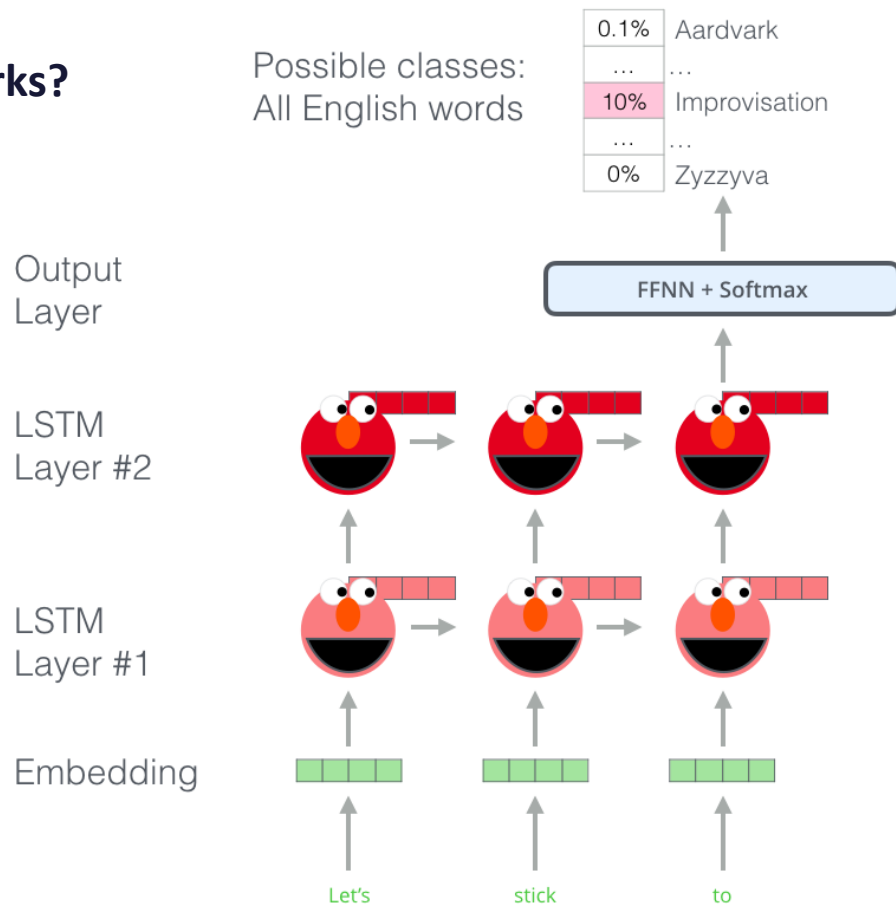
Words to embed



Natural Language Processing

Transfer Learning Models: How ELMo works?

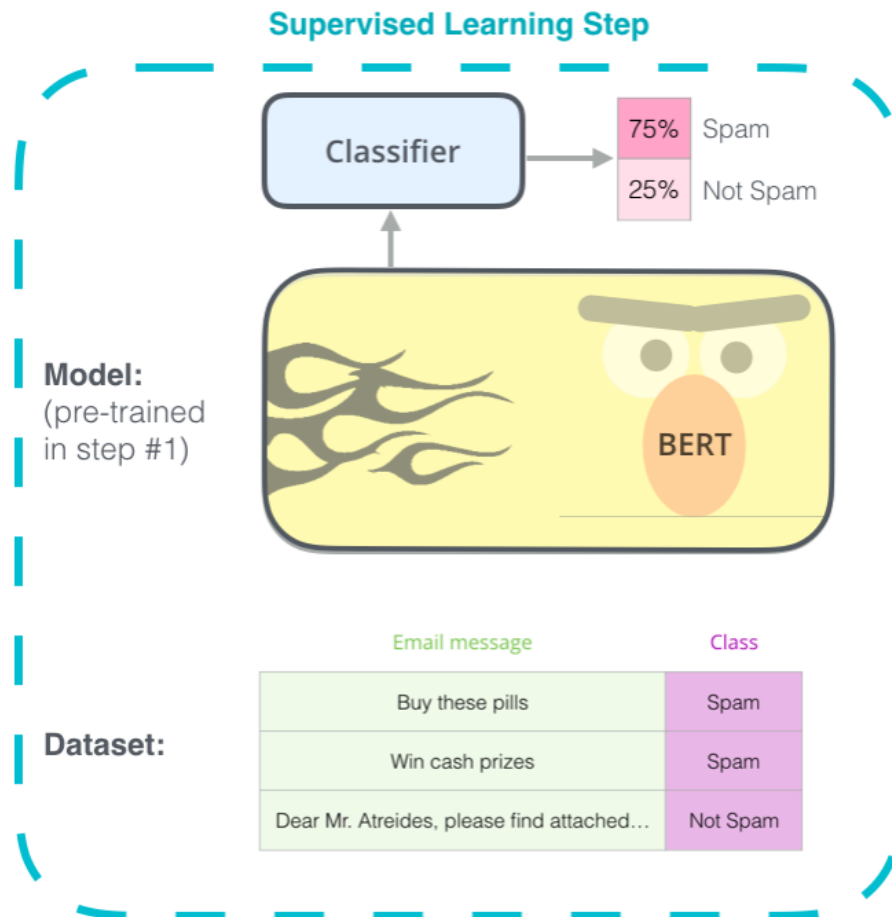
- Predecir la siguiente palabra
- Modelo pre-entrenado con redes recurrentes LSTM con datos no etiquetados (7000 libros).
- Fine tuning no-supervisado y si supervisado.



Natural Language Processing

Transfer Learning Models: BERT

- Entrenado con datos no etiquetados
- Puede utilizarse como feature-extractor o para realizar fine-tuning sobre él
- Estado del arte en diversos problemas de NLP.



Natural Language Processing

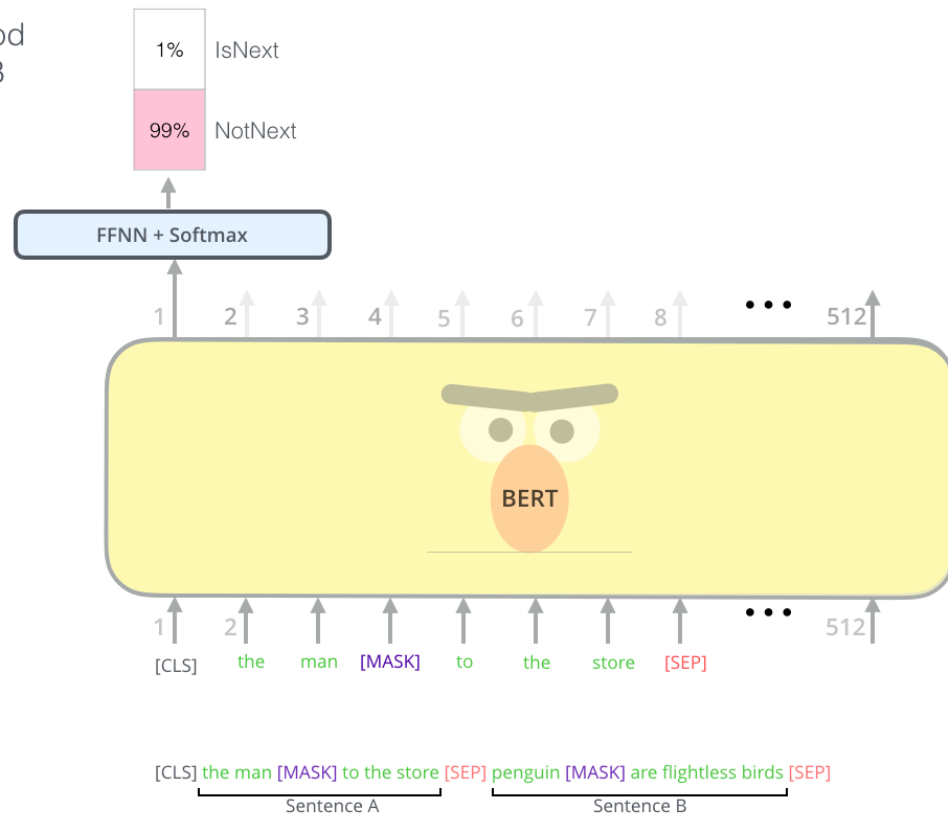
Transfer Learning Models: How BERT works?

- El primer objetivo es predecir dos palabras aleatorias dentro de la frase según su contexto
- El segundo objetivo es predecir si la segunda frase es la siguiente a la primera (binario)

Predict likelihood that sentence B belongs after sentence A

Tokenized Input

Input



Natural Language Processing

Transfer Learning Models: How BERT works?

[CLS] the man [MASK] to the store [SEP] penguin [MASK] are flightless birds [SEP]

└────────────────────────────────┘ └────────────────────────────────┘

Sentence A Sentence B

Natural Language Processing

Transfer Learning Models: ULM-FiT

- De los creadores de fastai
- Diseñado para hacer fine-tuning
- Buenos resultados aún con datasets de sólo 100 ejemplos etiquetados

Data Preparation

```
# Language model data
data_lm = TextLMDDataBunch.from_df(train_df = df_trn, valid_df = df_val, path = "")

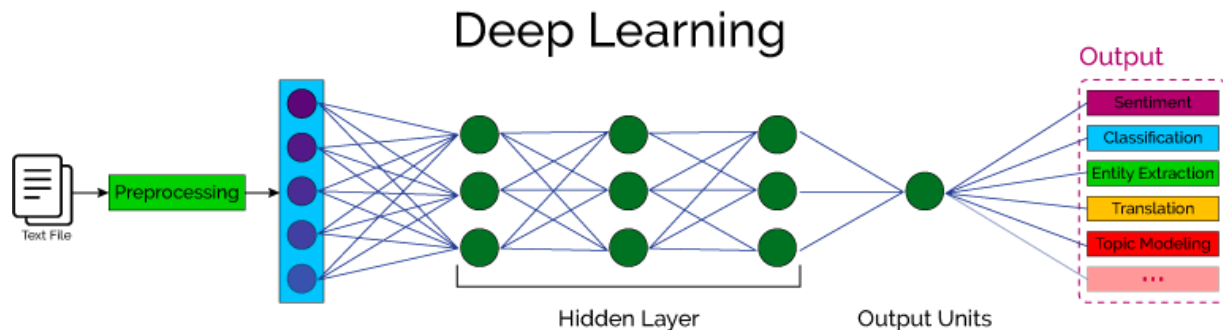
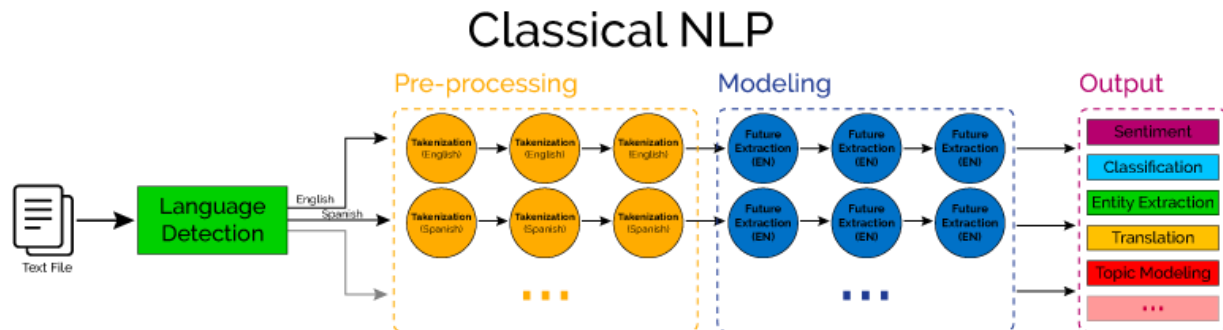
# Classifier model data
data_clas = TextClasDataBunch.from_df(path = "", train_df = df_trn, valid_df = df_val, vocab=data_lm.train_ds.vocab, bs=32)
```

```
learn = language_model_learner(data_lm, pretrained_model=URLs.WT103, drop_mult=0.7)
```

```
# train the learner object
learn.fit_one_cycle(1, 1e-2)
```

Natural Language Processing

Summary



Embeddings

BERT / ELMO / ULM-FiT

Your
Task

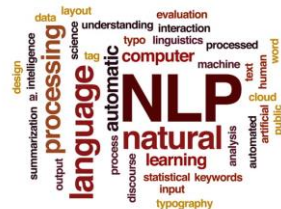
Natural Language Processing

Transfer Learning in NLP. Summary

- Word2vec > OneHot
- Modelos nuevos que detectan relaciones complejas (BERT, ELMo, ULM-Fit..)
- Disponible en distintos lenguajes, incluido el castellano.
- Modelos desde 2012-2013, es el momento de aprender Deep Learning + NLP

Recursos Transfer Learning NLP

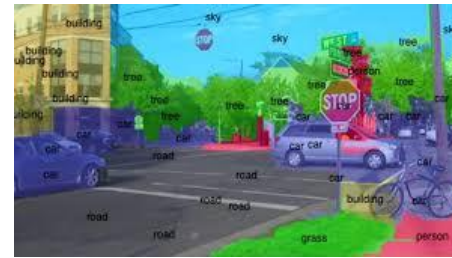
Transfer Learning as first iteration



Model	Resource	Paper
<i>fasttext</i>	https://fasttext.cc/docs/en/english-vectors.html	https://arxiv.org/abs/1607.04606
<i>GloVe</i>	https://nlp.stanford.edu/projects/glove/	https://www.aclweb.org/anthology/D14-1162
<i>word2vec</i>	https://code.google.com/archive/p/word2vec/	https://arxiv.org/pdf/1310.4546.pdf
<i>ELMo</i>	https://github.com/allenai/allennlp/blob/master/tutorials/how_to_elmo.md	https://arxiv.org/abs/1802.05365
<i>BERT</i>	https://github.com/google-research/bert	https://arxiv.org/abs/1810.04805
<i>ULMFit</i>	http://nlp.fast.ai/classification/2018/05/15/introducing-ulmfit.html	https://arxiv.org/abs/1801.06146
<i>USE</i>	https://tfhub.dev/google/universal-sentence-encoder/2	https://arxiv.org/abs/1803.11175

Recursos Transfer Learning Computer Vision

Transfer Learning as first iteration



Model	Resource	Paper
<i>Xception</i>	https://keras.io/applications/#xception	https://arxiv.org/abs/1610.02357
<i>VGG19</i>	https://keras.io/applications/#vgg19	https://arxiv.org/abs/1409.1556
<i>ResNet</i>	https://keras.io/applications/#resnet	https://arxiv.org/abs/1512.03385
<i>InceptionV3</i>	https://keras.io/applications/#inceptionv3	https://arxiv.org/abs/1409.4842
...



Keras Application

Recursos

<https://www.nytimes.com/2018/11/18/technology/artificial-intelligence-language.htm>

<https://medium.com/@14prakash/transfer-learning-using-keras-d804b2e04ef8>

<https://www.analyticsvidhya.com/blog/2018/11/tutorial-text-classification-ulmfit-fastai-library/>

https://blog.insightdatascience.com/using-bert-for-state-of-the-art-pre-training-for-natural-language-processing-1d87142c29e7?fbclid=IwAR2fgnyvpDtZKxBFZiAmojejf4MT8RnG_hdlzZcS0O0HZRGxWxYWHX7baQE

<https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>

<http://ruder.io/nlp-imagenet/>

<https://jalammar.github.io/illustrated-bert/>

<https://medium.com/owkin/transfer-learning-and-the-rise-of-collaborative-artificial-intelligence-41f9e2950657>

<https://keras.io/applications/>

<http://nlp.fast.ai/>

Contacto y preguntas

Pablo Vargas Ibarra
pablovargasibarra12@gmail.com

Manuel López Sheriff
manuel.lopez.sheriff@gmail.com

