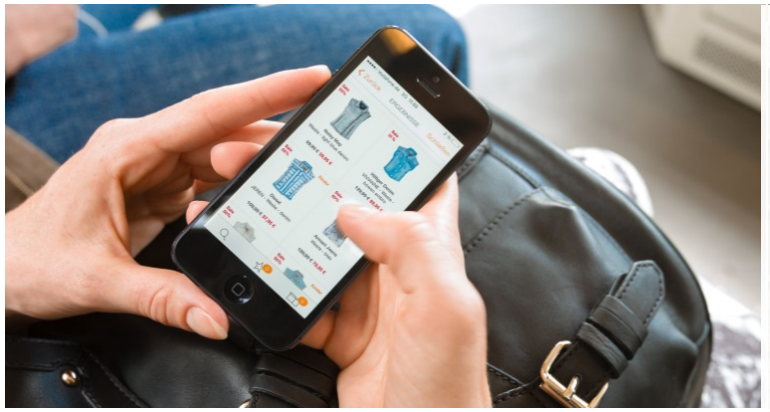


zalando

Recurrent Neural Networks (RNNs) for NLP

MACHINE LEARNING MEETUP

DR. ANA PELETEIRO RAMALLO

29-05-2017

TABLE OF CONTENTS

zalando

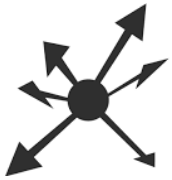
ZALANDO

Zalando is **the largest fashion** platform in Europe.

Zalando Tech employs **1600+** people in tech.

Our purpose: **to deliver** award-winning, best-in-class **shopping**

Experiences to our **+20 million** customers.



Radical agility:
- Purpose, autonomy and mastery

zalando

FASHION INSIGHTS CENTER

- Zalando Fashion Insights Centre was founded with the aim of understanding fashion through technology.
- R&D work to organize the world's fashion knowledge.
- We work with one of the richest datasets in eCommerce; products, profiles, customers, purchasing and returns history, online behavior, Web information and social media data.
- Three main teams:
  - Smart Product Platform
  - Customer Data Platform
  - Fashion Content Platform

zalando

NLP TEAM

- Not aiming to replace an stylist, but why not to help him/her?
- What is trending? What will people wear next year?
- Data driven decisions for the company.
- Fashion text is very complex, challenges!
  - Informality
  - Stylistic variance
  - Rich domain-specific language

FASHION TEXT EXAMPLES

The new crop of fall bags is a sumptuous parade of rich jewel tones, from Alexander Wang' s lush, matte emerald to Lanvin' s decorated sapphire, from Jason Wu' s gleaming garnet to Marc Jacob' s quilted topaz, and finally Judith Leiber' s bedazzled clutch, bursting with actual stones of amethyst, adventurine, sodalite, and Austrian crystals. The styles cover as wide a range as the palette.

Dries Van Noten' s fall 2015 collection, unveiled yesterday at the Hôtel de Ville in central Paris, was an Asian - inspired feast , from imperial brocade coats with Mongolian fur collars and khaki cotton duck trousers and work shirts with militant simplicity to dragon-embroidered bomber jackets and bead-embellished scenes of a rural Chinese village on voluminous skirts and delicate silks.

DEEP LEARNING FOR NLP

- Deep learning is having a transformative impact in many areas where machine learning has been applied.
- NLP was somewhat behind other fields in terms of adopting deep learning for applications.
- However, this has changed over the last few years, thanks to the use of RNNs, specifically LSTMs, as well as word embeddings.
- Distinct areas in which deep learning can be beneficial for NLP tasks, such as in named entity recognition, machine translation and language modelling, parsing, chunking, POS tagging, amongst others.

zalando

WORD EMBEDDINGS

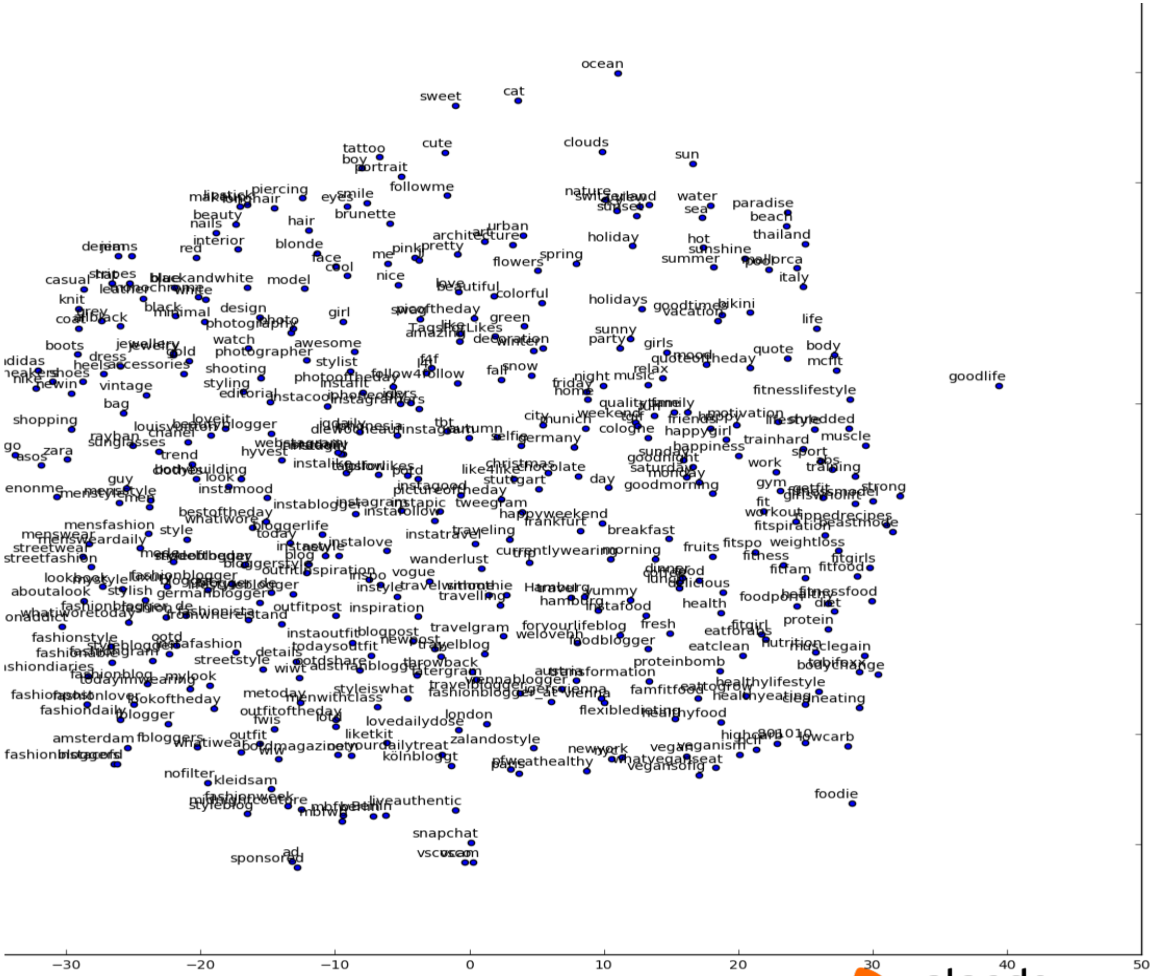
- Representing as ids.
  - Encodings are arbitrary.
  - No information about the relationship between words.
  - Data sparsity.
- Better representation for words.
  - Words in a continuous vector space where semantically similar words are mapped to nearby points.
  - Learn dense embedding vectors.
  - Skip-gram and CBOW
    - CBOW predicts target words from the context. E.g., Zalando ?? Talk
    - Skip-gram predicts source context-words from the target words. E.g., ?? Meetup ??
- Standard preprocessing step for NLP.
- Used also as a feature in supervised approaches (e.g., clustering)
- Several parameters we can experiment with, e.g., the size of the word embedding or the context window.

https://www.tensorflow.org/tutorials/word2vec

zalando

T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)

- Dimensionality reduction for high dimensional data.
- Very well suited for visualization of high dimensional datasets.
- Models each high-dimensional object by a two- or three-dimensional point in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points



zalando

Interesting relationships in the data:

E.g., Closest word to Manolo is Blahnik

zalando

RECURRENT NEURAL NETWORKS

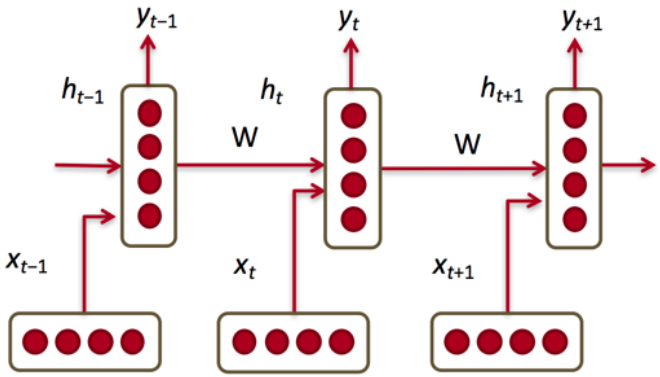
Why not basic Deep Nets?

- Traditional neural networks do not use information from the past, each entry is independent.
- This is fine for several applications, such as classifying images.
- However, several applications, such as video, or language modelling, rely on what has happened in the past to predict the future.
- Recurrent Neural Networks (RNN) are capable of conditioning the model on previous words in the corpus.

Language models

- Language models compute the probability of occurrence of a number of words in a particular sequence.
- First, it allows us to score arbitrary sentences based on how likely they are to occur in the real world (useful for machine translation).
- A language model allows us to generate new text.
- Problem with traditional approaches: only takes a fixed window into account.
- Recurrent neural networks do not use limited size of context.

RNNs

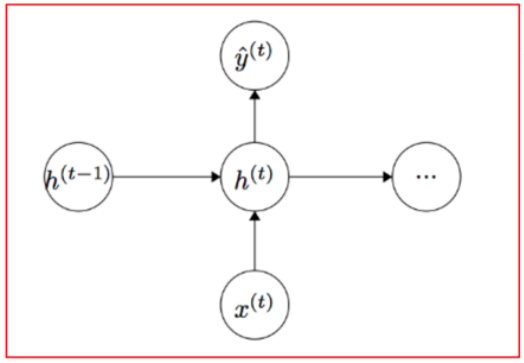


Richard Socher 4/21/16

$$x_1, \ldots, x_{t-1}, x_t, x_{t+1}, \ldots, x_T$$

$$h_t = \sigma\left(W^{(hh)} h_{t-1} + W^{(hx)} x_{[t]}\right)$$

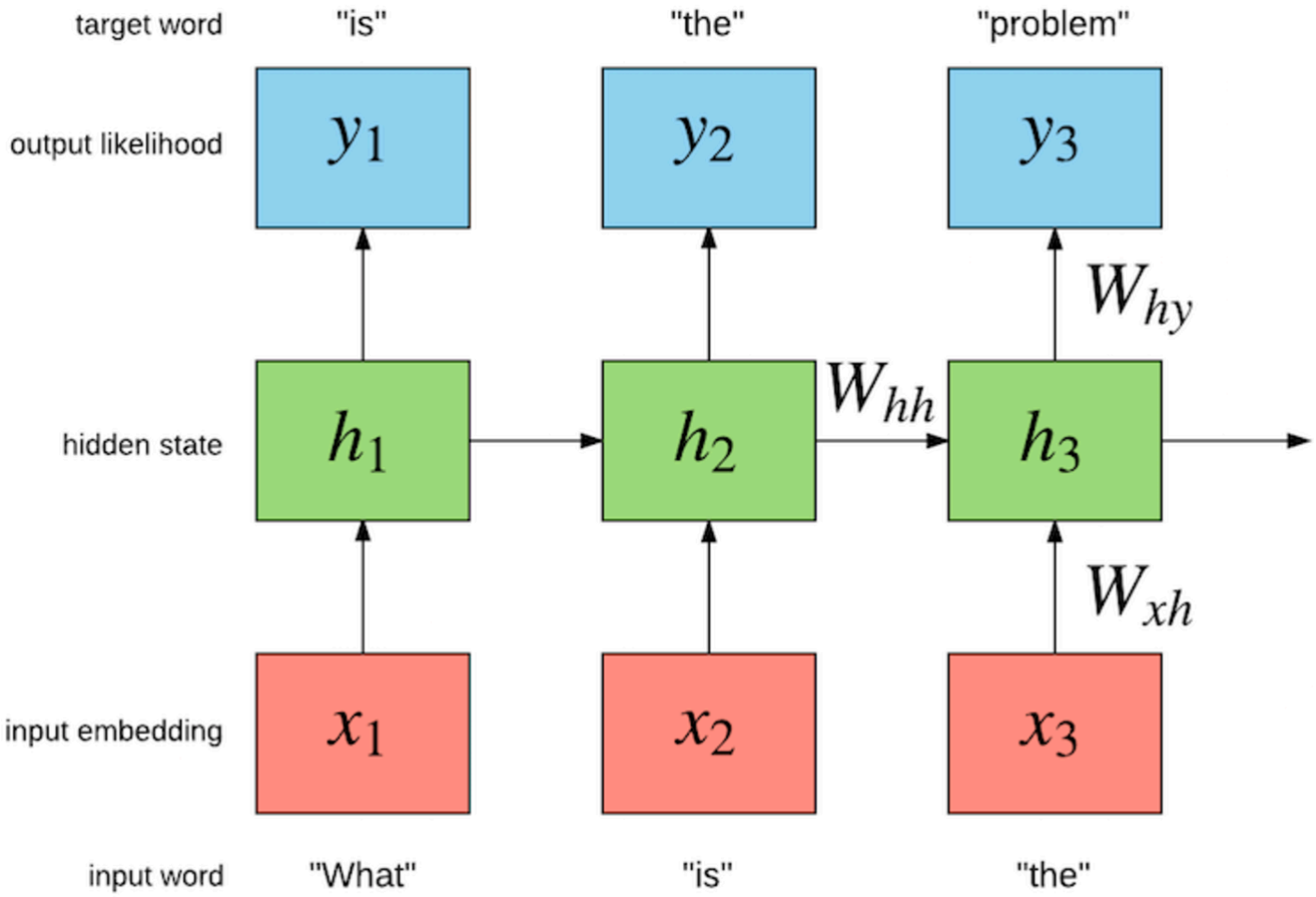
$$\hat{y}_t = \text{softmax}\left(W^{(S)} h_t\right)$$



- Make use of sequential information.
- Output is dependent on the previous information.
- RNN shares the same parameter W for each step, so less parameters we need to learn.

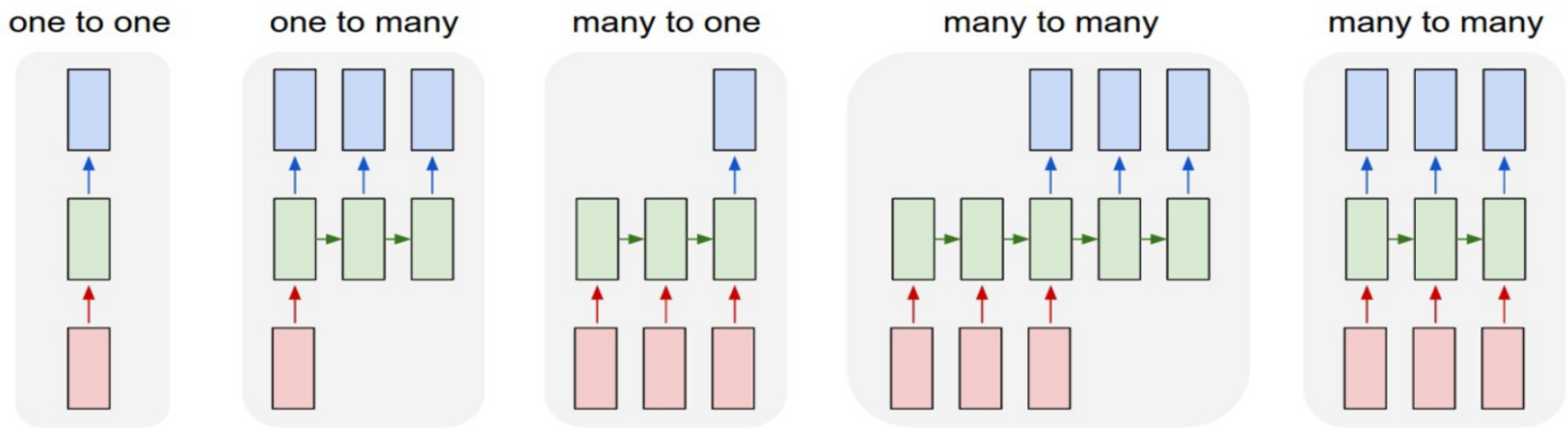
http://cs224d.stanford.edu/lectures/CS224d-Lecture8.pdf

zalando



http://torch.ch/blog/2016/07/25/nce.html

RNN architectures



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

zalando

RNNs (II)

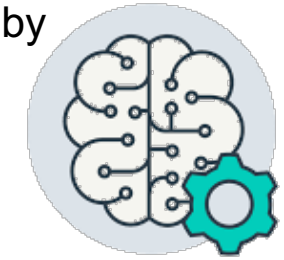
- In theory, RNNs are absolutely capable of handling such long-term dependencies. Practice is ”a bit” different.
- Parameters are shared by all time steps in the network, the gradient at each output depends not only on the calculations of the current time step, but also the previous time steps.
- Exploding gradients:
  - Easier to spot.
  - Clip the gradient to a maximum
- Vanishing gradients:
  - Harder to identify
  - Initialization of the matrix to identity matrix
  - Relus instead of sigmoid

Long Short Term Memory (LSTMs)

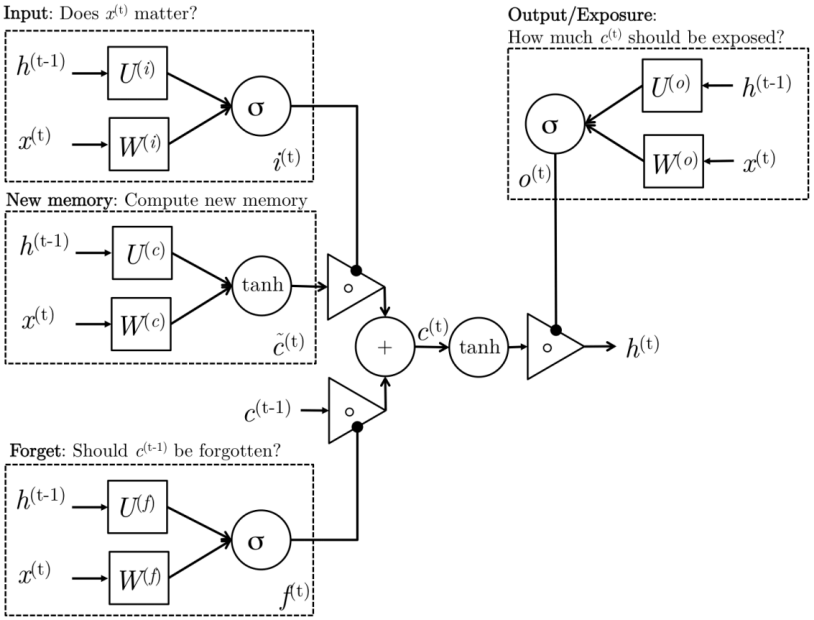
- In theory, RNNs can handle of handling such long-term dependencies.

The oversized mannish coats looked positively edible over the bun-skimming dresses while combined with novelty knitwear such as punk-like fisherman's sweaters. As other look, the ballet pink Elizabeth and James jacket provides a cozy cocoon for the 20-year-old to top off her ensemble of a T-shirt and Parker Smith jeans. But I have to admit that my favorite is the bun-skimming dresses with the ??

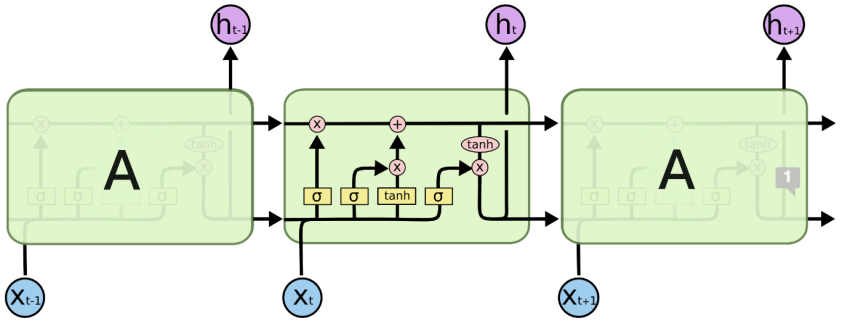
- However, in reality, they cannot.
- LSTMs avoid the long-term dependency problem.
- Remove or add information to the cell state, carefully regulated by structures called gates.
- Gates are a way to optionally let information through.



LSTMs



$$
\begin{aligned}
i_t &= \sigma(W^{(i)}x_t + U^{(i)}h_{t-1}) && \text{(Input gate)} \\
f_t &= \sigma(W^{(f)}x_t + U^{(f)}h_{t-1}) && \text{(Forget gate)} \\
o_t &= \sigma(W^{(o)}x_t + U^{(o)}h_{t-1}) && \text{(Output/Exposure gate)} \\
\tilde{c}_t &= \tanh(W^{(c)}x_t + U^{(c)}h_{t-1}) && \text{(New memory cell)} \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t && \text{(Final memory cell)} \\
h_t &= o_t \circ \tanh(c_t)
\end{aligned}
$$



http://colah.github.io/posts/2015-08-Understanding-LSTMs/

http://cs224d.stanford.edu/lecture_notes/notes4.pdf

zalando

HACK WEEK PROJECT

- Our annual, week-long celebration of open innovation and experimentation, where technologists are free to work on inspiring, inventive new projects for the business.
- We were working on different Deep Learning problems, with the available data that we have.
- We won the best software development award
- Want to know some more? Read here and here

zalando

Language modelling example

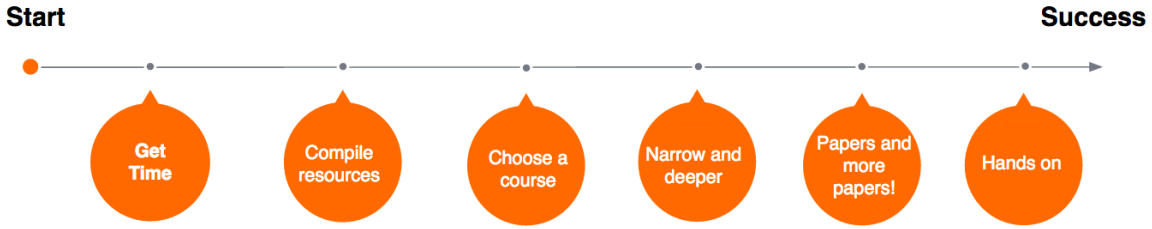
Fake or real?

There was something so very interesting about the idea, these 1650 nipped-in jackets with these Deauville-y cropped trousers and these sun hats.

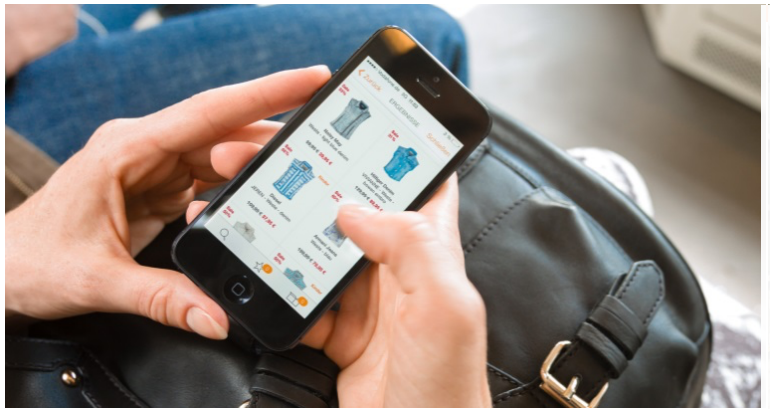
Erdem Moralioglu tells WWD about his Spring 2017 collection. They stand out and if you aren't easy on wearing heels for a scalloped cowgirl look for toting them over.

UPSKILLING AND CONSIDERATIONS IN DATA SCIENCE DELIVERY

- Have a look at our blogpost **Sapphire Deep Learning Upskilling**!
- Compile resources.
- Choose a course
  - Deep Learning by Google.
- Narrow NLP
  - NLP Stanford classes.
    - Lectures
    - Other related materials
- Read papers, papers and more papers.
- Get hands on



zalando



zalando

THANKS!

ana.peleteiro@zalando.com

@PeleteiroAna

29-05-2017