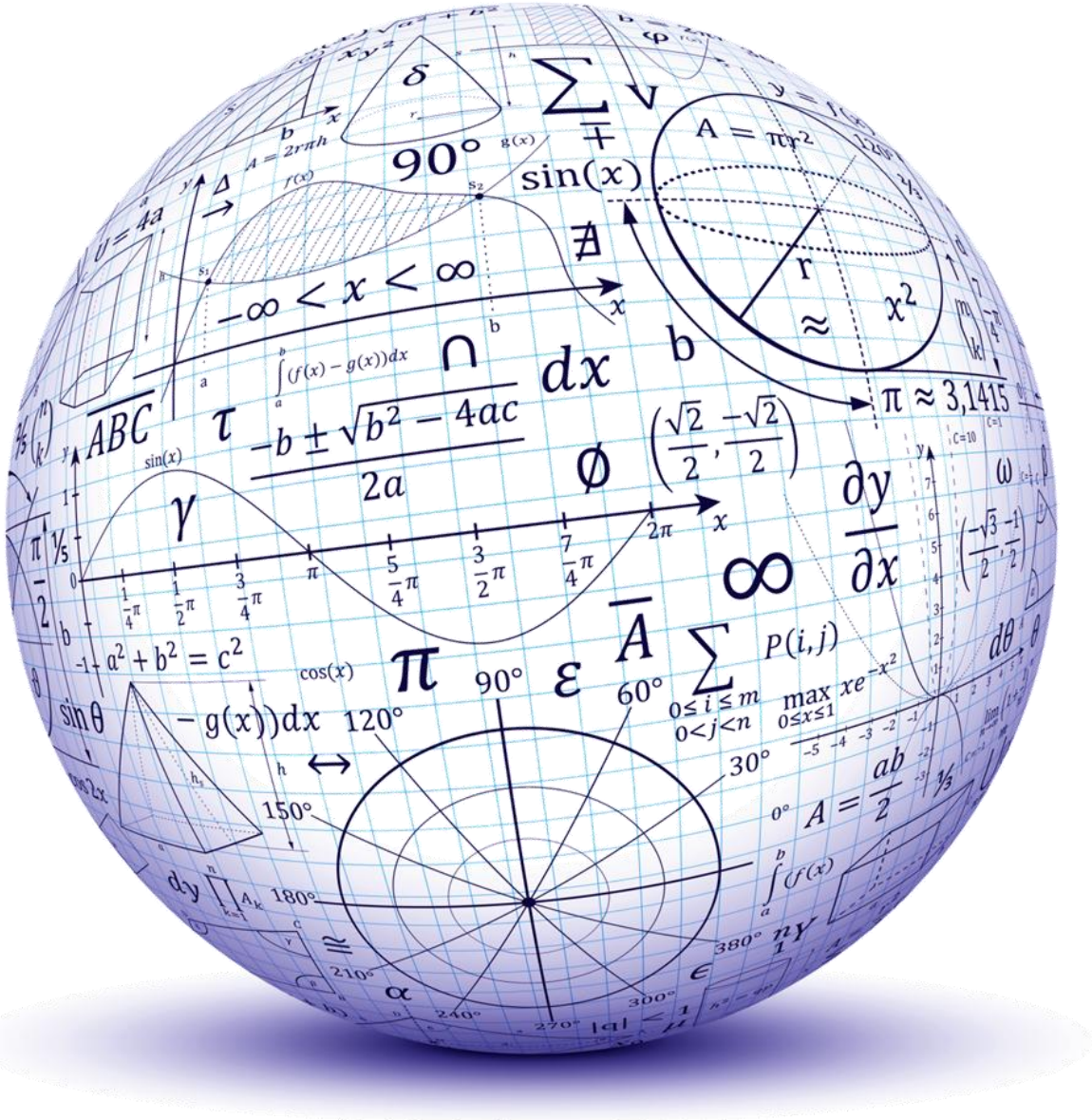
RES ACADEMY
INNOVATION & RESEARCH



Res Academy

Reti Neurali
Named Entity Recognition

4 maggio 2017

RES ACADEMY
INNOVATION & RESEARCH

Sommario

- Gruppo RES
- RES Institute for Data Science (RIDS)
- Neural Network e back propagation, esempio con metodo del gradiente
- Gradiente stocastico
- Back propagation, esempio con metodo del gradiente stocastico
- Named Entity Recognition
- NER, esempio con metodo del gradiente stocastico
- Altri metodi per il problema NER
- Conclusioni

Property of RES Academy

RES ACADEMY
INNOVATION & RESEARCH

Il Gruppo RES

Chi siamo

Breve storia dalla fondazione di RES ad oggi:

- RES nasce nel 1987 e si sviluppa in un mercato di nicchia dell'IT in forte crescita grazie all'avvento dei DB relazionali.
- Negli anni '90 RES investe nello sviluppo di tecnologie nell'ambito dell'ottimizzazione dei Sistemi Informativi di grandi dimensioni: banche, assicurazioni e grandi imprese.
- Nel decennio dal '95 al 2005, RES con i suoi prodotti di documentazione e reenginering nell'ambito dei sistemi Mainframe IBM, diventa leader sul mercato italiano.
- Negli anni dal 2006 al 2014, nonostante la crisi finanziaria mondiale, RES riesce a rinnovare completamente la sua offerta ricoprendo tutte le piattaforme.
- Oggi RES, rappresenta un Gruppo di 6 imprese con oltre 7 milioni di euro di fatturato, più di 100 collaboratori ed è diffusa a livello internazionale in Europa e nelle Americhe

Property of RES Academy

Il Gruppo RES

La crescita e le prospettive

- RES nel 2017, grazie ad una solida base di partenza e ad una attenta pianificazione della propria offerta di mercato, rappresenta un gruppo in crescita con importanti progetti di sviluppo già in corso o pronti ad essere avviati su fronti diversificati

  - ✓ RES PGB – il potenziamento dei servizi IT
    acquisire una nuova fetta di mercato complementare al core business di RES

  - ✓ RES Academy for Research – la "fucina delle idee"
    investire in formazione, ricerca e sperimentazione tramite la cooperazione con le Università scientifiche, con particolare riferimento alle facoltà di Ingegneria, Matematica e Statistica

  - ✓ RES Academy for Startup – come tradurre in impresa idee e progetti innovativi
    tramite un supporto a 360°, fornire a giovani laureandi e neolaureati l'opportunità di fare impresa per realizzare i propri progetti

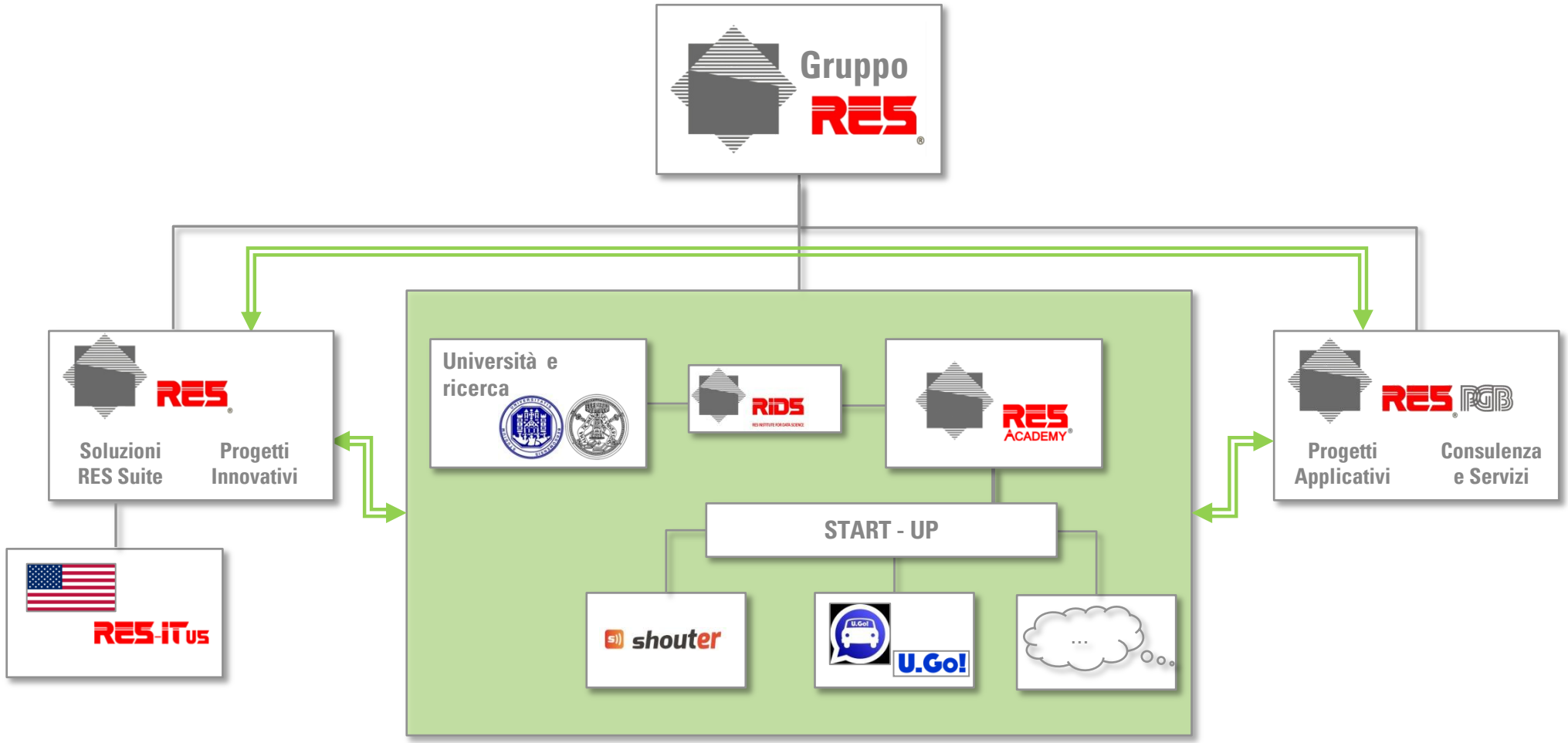
  - ✓ RES Institute for Data Science (RIDS) – il laboratorio dell'Università di Pavia per la realizzazione di progetti consulenziali di Data Science, in supporto alla ricerca universitaria

- Una linea guida comune: favorire le "correlazioni virtuose" fra le varie iniziative intra-gruppo

Property of RES Academy

RES ACADEMY
INNOVATION & RESEARCH

RES – L'offerta di Gruppo



Gruppo RES – il Mercato



Settori

Banche

Assicurazioni

Società Finanziarie

Servizi

Pubblica Amministrazione

Grande distribuzione

Università

Industria

Utility

L'offerta "Res Academy"

Vision

L'offerta Res Academy è nata per rispondere alla sempre crescente esigenza di previsioni strategiche da parte del mercato.

Utilizzare la ricerca in ambito statistico e di modellazione matematica, con la tecnologia fornita da RES è la strategia più efficace per riuscire a soddisfare le esigenze di innovazione del cliente

Property of RES Academy

Il Progetto “Res Academy”

Le componenti

Competenze generali
Strategia IT
Implementazione SI
Sviluppo SW
Progettazione

Competenze Scientifiche
Modellazione matematica
Analisi statistica e previsionale
Machine learning

RIDS(Res Institute for Data Science)
Laboratorio dell’ Università di Pavia
Progetti consulenziali per le aziende
Progetti di ricerca applicata per le aziende

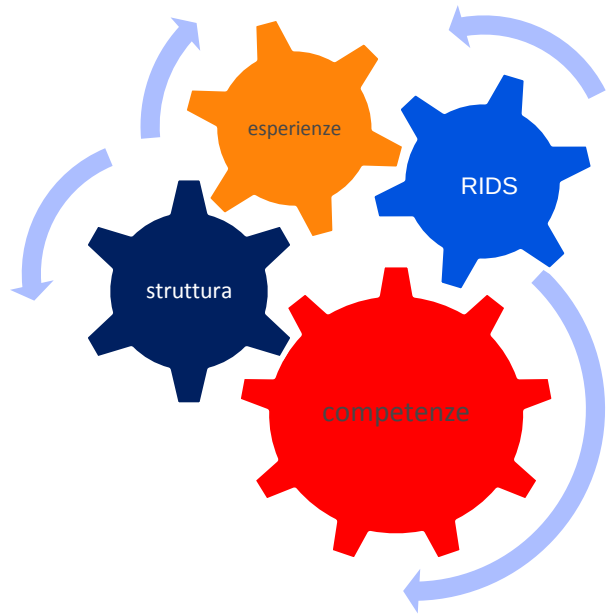
Esperienze
Business Case
Studi in fase di applicazione

Struttura
Internamente:

Data scientist (matematici applicati, statistici computazionali)
Ingegneri specializzati nel trasferimento tecnologico
sviluppatori e matematici
Organizzazione "agile programming" con metodologie Scrum e Kanban

Esternamente:

Gruppo di ricerca del prof. Paolo Giudici
Gruppo di ricerca Risk Data Science della prof. Silvia Figini
Stagisti dell’Università di Pavia, indirizzi di Matematica, Economia, Scienze Politiche, Filosofia e analisi del linguaggio



Property of RES Academy

Il laboratorio RES Institute for Data Science (RIDS)

La novità

Come si può consentire alle aziende italiane l'accesso a ricerche di altissimo livello?
*Per la Data Science, la risposta è il **RIDS**.*

Dalla collaborazione con il mondo accademico nasce la Joint Venture con L'Università di Pavia con la quale vengono portati a realizzazione progetti di prodotto e di ricerca applicata per le aziende che ne volessero usufruire.

Alcuni dei temi già trattati:

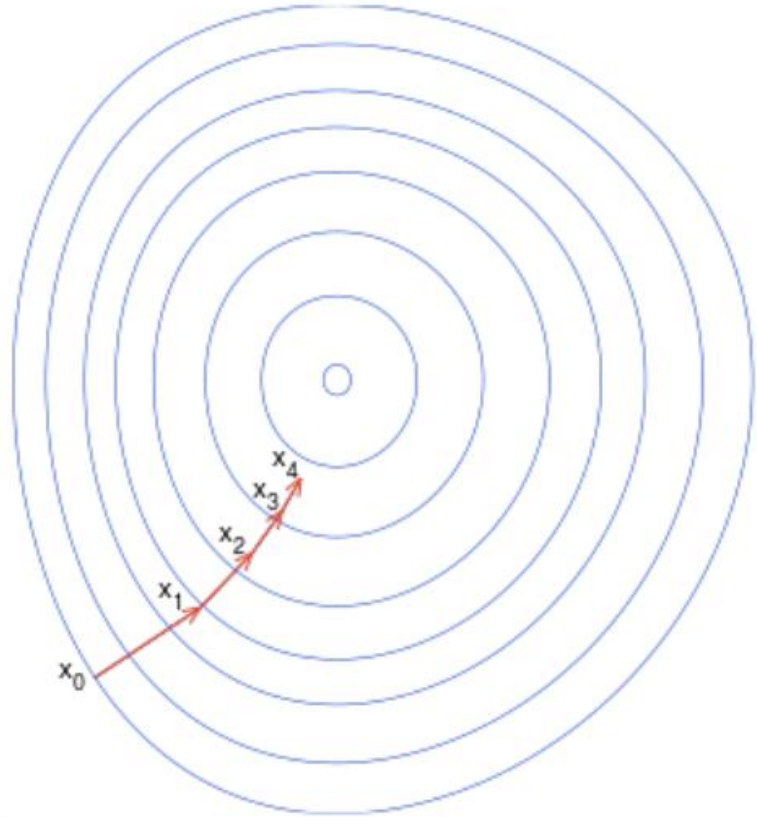
- Anticorruzione
- Antiriciclaggio
- Credit Risk
- Finanziamento al terrorismo e previsione di attacchi
- Rischio reputazionale
- ….

Property of RES Academy

Gradient Descent

Characteristics

- Simplest and extremely popular
- Main Idea: take a **step proportional to the negative of the gradient**
- Easy to implement
- Each iteration is relatively cheap
- **Can be slow** to converge



Property of RES Academy

Example: Linear Regression

- Optimization problem: $$\min_w ||Xw - y||_2$$
- Closed form solution: $$w^* = (X^\top X)^{-1} X^\top y$$
- Gradient update: $$w^+ = w - \frac{1}{m} \sum_i (x_i^\top w - y_i) x_i$$

Requires an entire pass through the data!

Property of RES Academy



RES ACADEMY
INNOVATION & RESEARCH

Tackling Compute Problems: Scaling to Large n

Main challenges

- **Streaming** implementation
- **Parallelize** your batch algorithm
- Aggressively subsample the data
- Change algorithm or training method
  - Optimization is a surrogate for learning
  - Trade-off weaker optimization with more data

Property of RES Academy

Stochastic Optimization

Motivation

- Information is redundant amongst samples
- Sufficient samples means we can afford more frequent, noisy updates
- Never-ending stream means we should not wait for all data
- Tracking non-stationary data means that the target is moving

Property of RES Academy

Stochastic Optimization

Characteristics

- Idea: Estimate function and **gradient from a small, current subsample** of your data and with enough iterations and data, you will converge to the true minimum
- Pro: Better for large datasets and often faster convergence
- Con: Hard to reach high accuracy
- Con: Best classical methods can't handle stochastic approximation
- Con: Theoretical definitions for convergence not as well defined

Property of RES Academy

Stochastic Gradient Descent (SGD)

Weight update

- Randomized gradient estimate to minimize the function using a single randomly picked example

$$\text{Instead of } \nabla f, \text{use } \tilde{\nabla} f, \text{where } E[\tilde{\nabla} f] = \nabla f$$

The resulting update is of the form:

$$w^{+} = w - \gamma \nabla_w L(f_w(x_i, y_i))$$

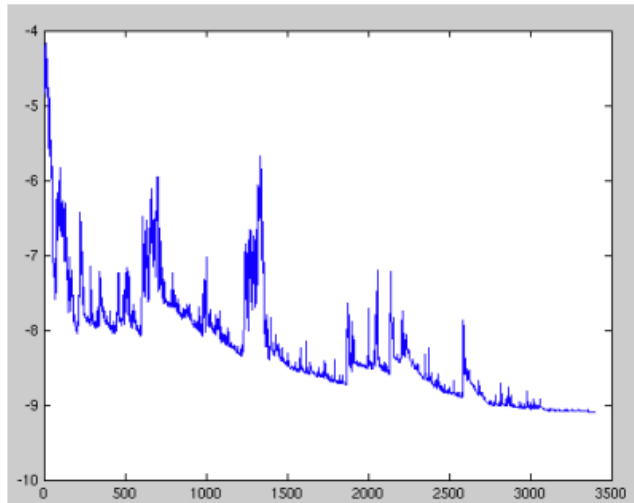
Although random noise is introduced, it **behaves like gradient descent in its expectation**

Property of RES Academy

SGD Algorithm

Pseudocode

Randomly initialize parameter w and learning rate γ
while *Not Converged* **do**
  Randomly shuffle examples in training set
  for $i = 1, \cdots, N$ **do**
    $w^+ = w - \gamma \nabla_w L(f_w(x_i, y_i))$
  end
end



Property of RES Academy



RES ACADEMY
INNOVATION & RESEARCH

The Benefits of SGD

- Gradient is easy to calculate (“instantaneous”)
- Less prone to local minima
- Small memory footprint
- Get to a reasonable solution quickly
- Works for non-stationary environments as well as online settings
- Can be used for more complex models and error surfaces

Property of RES Academy

Mini-batch Stochastic Gradient Descent

Characteristics

- Rather than using a single point, use a random subset where the size is less than the original data size

$$w^+ = w - \gamma \frac{1}{|S_k|} \sum_{i \in S_k} \nabla_w L(f_w(x_i, y_i)), \text{where } S_k \subseteq [n]$$

Like the single random sample, the full gradient is approximated via an unbiased noisy estimate

Random subset reduces the variance by a factor of 1/|Sk|, but is also |Sk| times more expensive

Property of RES Academy

Example: Regularized Logistic Regression

- Optimization problem:

$$\min \frac{1}{n} \sum_i \left(-y_i x_i^\top \beta + \log(1 + e^{x_i^\top \beta}) \right) + \frac{\lambda}{2} ||\beta||_2^2$$

- Gradient computation:

$$\nabla f(\beta) = \sum_i (y_i - p_i(\beta)) x_i + \lambda \beta$$

- Update costs:

  Batch: O(nd)

  Stochastic: O(d)

  Mini-batch: O(|Sk|d)

Batch is doable if n is moderate
but not when n is huge

Property of RES Academy

Asymptotic Analysis of GD and SGD

Bottou, L. (2012). Stochastic Gradient Descent Tricks.

Table 2. Asymptotic equivalents for various optimization algorithms: gradient descent (GD, eq. 2), second order gradient descent (2GD, eq. 3), stochastic gradient descent (SGD, eq. 4), and second order stochastic gradient descent (2SGD, eq. 5). Although they are the worst optimization algorithms, SGD and 2SGD achieve the fastest convergence speed on the expected risk. They differ only by constant factors not shown in this table, such as condition numbers and weight vector dimension.

	GD	2GD	SGD	2SGD
Time per iteration :	n	n	1	1
Iterations to accuracy ρ :	$\log \frac{1}{\rho}$	$\log\log \frac{1}{\rho}$	$1/\rho$	$1/\rho$
Time to accuracy ρ :	$n \log \frac{1}{\rho}$	$n \log\log \frac{1}{\rho}$	$1/\rho$	$1/\rho$
Time to excess error ε :	$\frac{1}{\varepsilon^{1/\alpha}} \log^2 \frac{1}{\varepsilon}$	$\frac{1}{\varepsilon^{1/\alpha}} \log \frac{1}{\varepsilon} \log\log \frac{1}{\varepsilon}$	$1/\varepsilon$	$1/\varepsilon$

Property of RES Academy

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:

  - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Person
Date
Location
Organi-
zation

Property of RES Academy



RES ACADEMY
INNOVATION & RESEARCH

Named Entity Recognition (NER)

- The uses:
  - Named entities can be indexed, linked off, etc.
  - Sentiment can be attributed to companies or products
  - A lot of IE relations are associations between named entities
  - For question answering, answers are often named entities.
- Concretely:
  - Many web pages tag various entities, with links to bio or topic pages, etc.
    - Reuters' OpenCalais, Evri, AlchemyAPI, Yahoo's Term Extraction, …
  - Apple/Google/Microsoft/… smart recognizers for document content

Property of RES Academy

Precision and Recall

- **Precision**: % of selected items that are correct
  Recall: % of correct items that are selected

	correct	not correct
selected	tp	fp
not selected	fn	tn

Property of RES Academy

Three standard approaches to NER

1. Hand-written regular expressions
   - Perhaps stacked
2. Using classifiers
   - Generative: Naïve Bayes
   - Discriminative: Maxent models
3. Sequence models
   - HMMs
   - CMMs/MEMMs
   - CRFs

Property of RES Academy

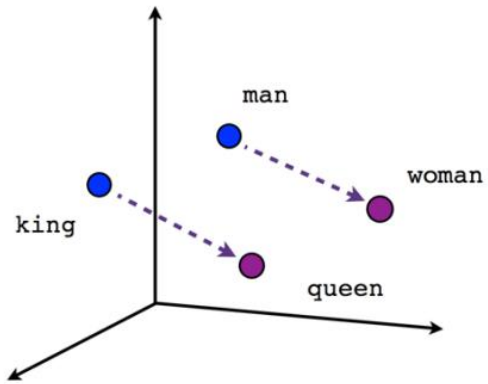
Training dataset for NER

Fred	PER
showed	O
Sue	PER
Mengqiu	PER
Huang	PER
‘s	O
new	O
painting	O

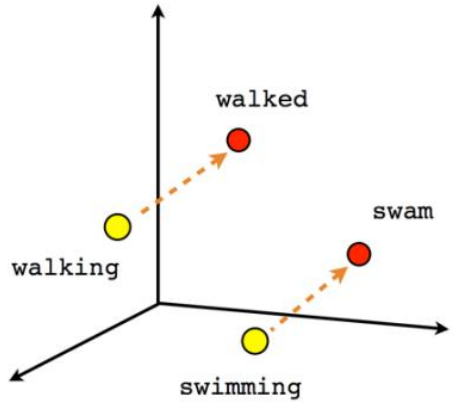
Property of RES Academy

Word Embeddings

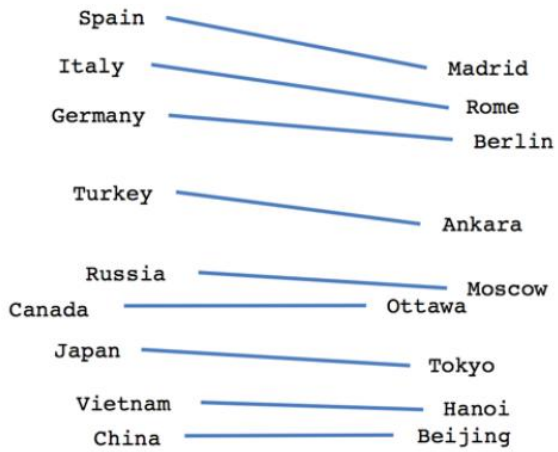
- Word embeddings map words to a reduced, continuous, semantic space where particular properties can be extracted by vector operations



Male-Female



Verb tense

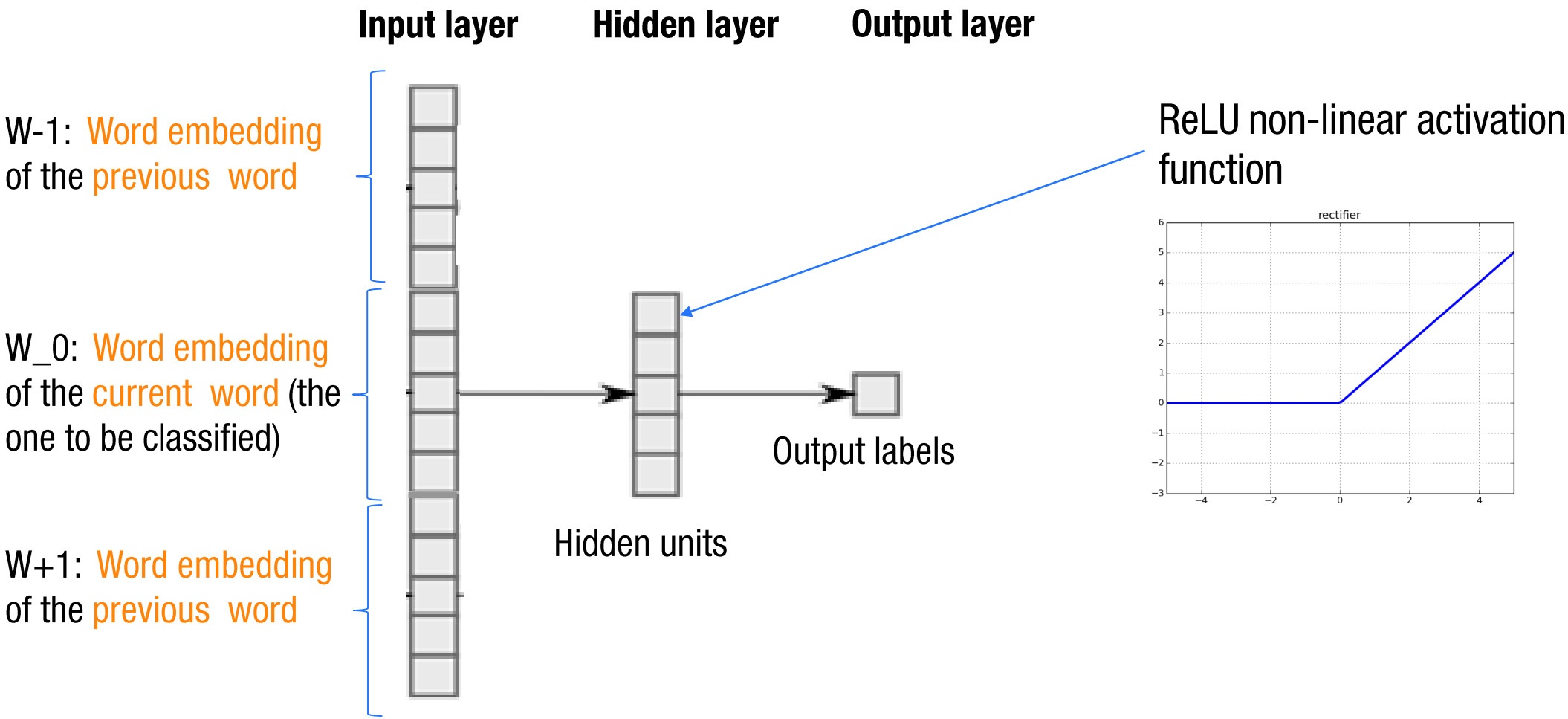


Country-Capital

- 200 - 800 dimensions depending on the task and the available corpus
- Words that are near in this space have similar semantic value
- Wiser space to operate classification task because of the reduced dimensionality and preserves the semantic value of words

Property of RES Academy

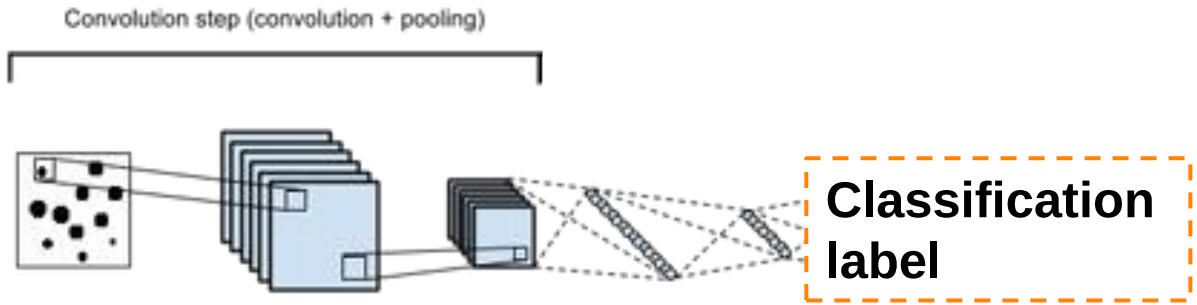
Word Embeddings



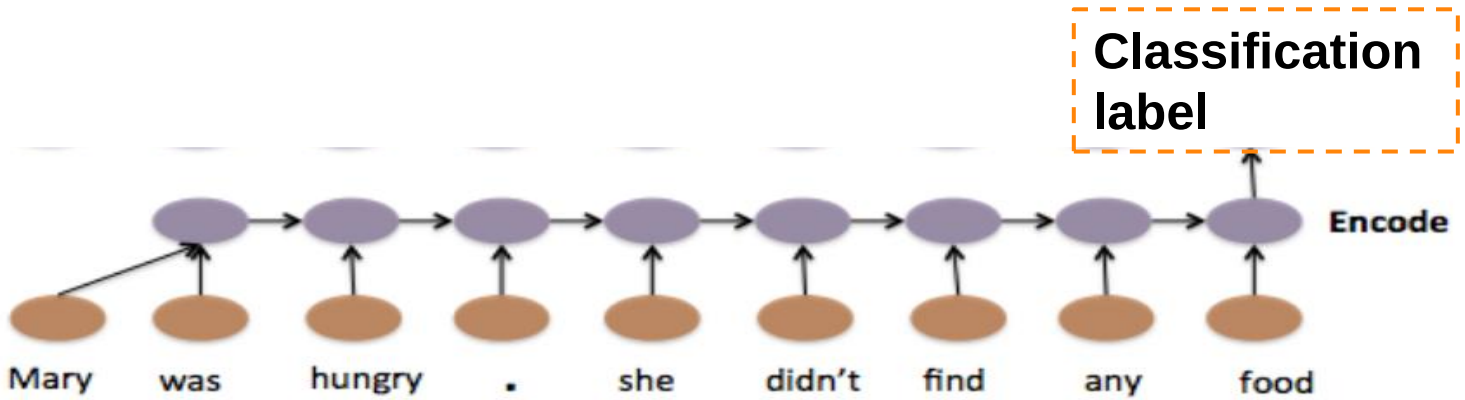
Property of RES Academy

Two of the most used neural network architectures today

- Convolutional Neural Network (classification)



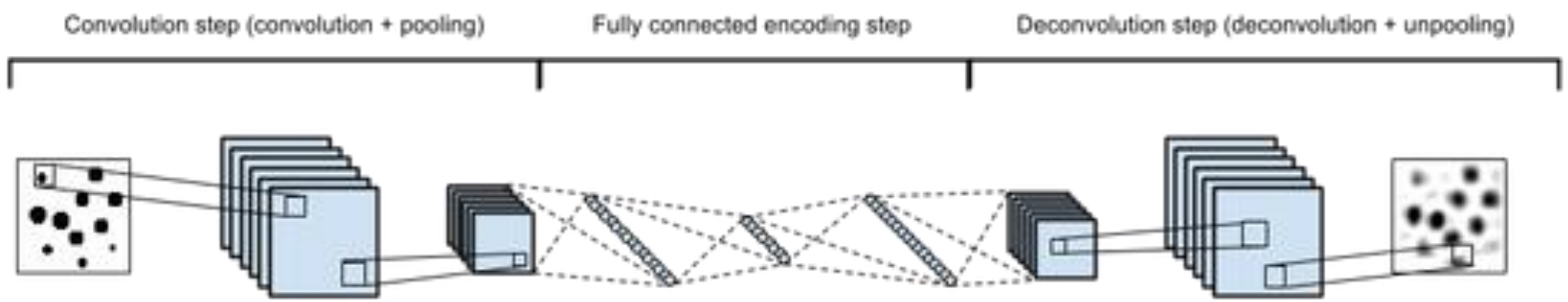
- Recurrent Neural Networks (classification)



Property of RES Academy

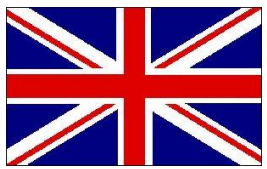
Two of the most used neural network architectures today

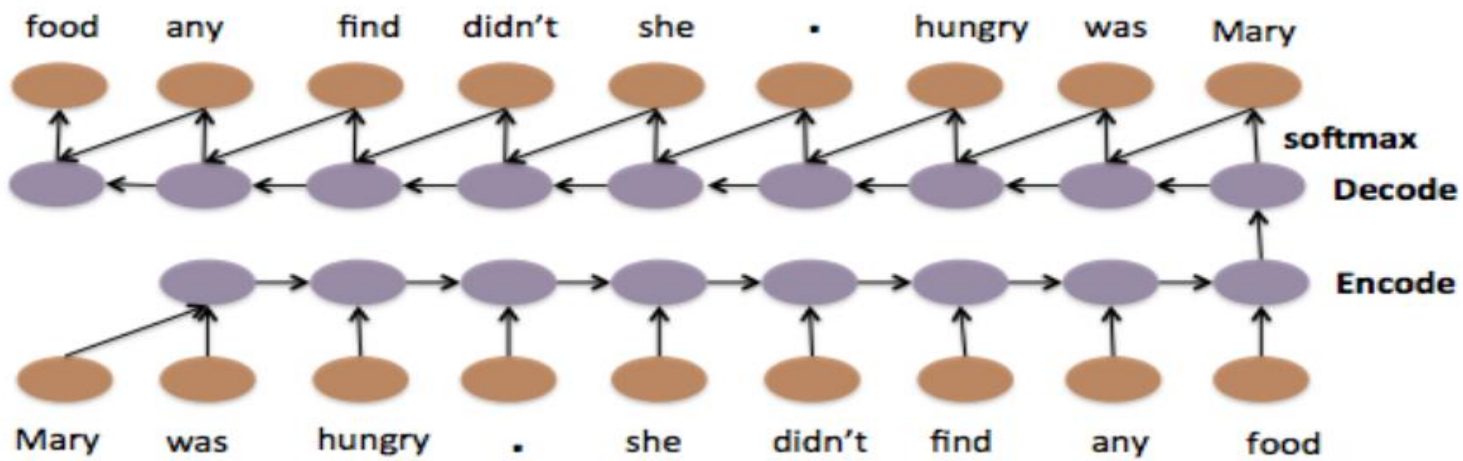
- Convolutional Neural Network (encoder + decoder)



- Recurrent Neural Networks (encoder + decoder)







Property of RES Academy

RES ACADEMY
INNOVATION & RESEARCH

Our Neural Network applications

Current projects

- Named Entity Recognition (NER)
- Part of Speech Tagging (POS)
- Sentiment analysis
- Text classification
- Automatic news analysis for bank stress identification

... Future projects

- Relation extraction and classification
- Event extraction and classification
- Part of Speech Tagging (POS)
- Similar text retrieval
- Text summarization
- Automatic image interpretation

Kaggle data challenges

- Quora duplicate question identification (top 19%)
- Leaf classification (top 12%)
- Fish recognition (top 42%)

Property of RES Academy

Ciclo di assunzioni 2017

Requisiti e profili

Chi cerchiamo?

Neolaureati e laureandi triennali e magistrali

Che competenze chiediamo?

Capacità di apprendimento, mentalità scientifica, competenze in matematica e statistica o informatica

Che posizione offriamo?

Stage formativo retribuito di 6 mesi, seguito da assunzione a tempo indeterminato

Che durata?

Lo stage formativo di 6 mesi è un investimento sulle persone, da parte di RES. Siamo interessati solo a candidati che prevedono di restare alla fine del periodo di stage

Che posizione offriamo?

Stage formativo retribuito di 6 mesi, seguito da assunzione a tempo indeterminato

Che contributo ci aspettiamo da un matematico?

Gli studenti di matematica hanno una inclinazione al ragionamento logico e forte facilità di apprendimento, imparano in fretta a programmare e posseggono competenze adatte per diventare Data Scientist

Come inviare CV?

www.res-it.com
→ Contatti → Lavora con noi

17 maggio
Career Day
Porte Aperte alle Imprese

Property of RES Academy

Federico Bonelli

Coordinamento Tecnico-Scientifico

federico.bonelli@res-it.com

M +39 333 4898541

Giancarlo Nicola

PhD researcher

giancarlo.nicola@res-it.com

M +39 334 3228037

RES ACADEMY
INNOVATION & RESEARCH

THANK YOU FOR YOUR ATTENTION