



ProApes

proapes11@gmail.com

Verbale Esterno 2020-04-03

Versione	1.0.0-1.10
Data approvazione	2020-04-06
Responsabile	Giacomo Piran
Redattori	Valentina Signor
Verificatori	Francesco Bari
Stato	Approvato
Lista distribuzione	<i>ProApes</i> <i>Prof. Tullio Vardanega</i> <i>Prof. Riccardo Cardin</i>
Uso	Esterno

Sommario

Resoconto dell'incontro esterno effettuato dal gruppo *ProApes* il giorno 2020-04-03 attraverso una videochiamata registrata mediante l'uso di *Skype*

Diario delle Modifiche

Versione	Data	Modifica	Autore	Ruolo
v1.0.0-0.4	2020-04-06	<i>Approvazione del documento</i>	Giacomo Piran	<i>Responsabile di Progetto</i>
v0.1.0-0.4	2020-04-05	<i>Revisione complessiva di coerenza e coesione (Verificatore: Francesco Bari)</i>	Valentina Signor	<i>Progettista</i>
v0.0.1-0.4	2020-04-03	<i>Stesura del documento (Verificatore: Francesco Bari)</i>	Valentina Signor	<i>Progettista</i>

Indice

1	Informazioni generali	3
2	Ordine del giorno	3
3	Resoconto	4
3.1	Scopo del <i>Machine Learning</i>	4
3.2	Origini del <i>Machine Learning</i>	4
3.3	La <i>maledizione della dimensionalità</i>	5
3.4	Confidenza delle previsioni ottenute	5
4	Riepilogo delle decisioni	6

1 Informazioni generali

- **Luogo:** - ;
- **Strumento:** *Skype*;
- **Data:** 2020-04-03;
- **Ora inizio:** 11.00;
- **Ora fine:** 12.30;
- **Segretario:** Valentina Signor;
- **Partecipanti:** il *Dott. Gregorio Piccoli* come rappresentante dell'azienda proponente *Zucchetti S.p.A.*, alcuni competitors e i componenti del gruppo *ProApes* anche se in modo asincrono:
 - Francesco Bari;
 - Igor Biolcati Rinaldi;
 - Fiammetta Cannavò;
 - Federico Carboni;
 - Alessandro Discalzi;
 - Giacomo Piran;
 - Valentina Signor.

2 Ordine del giorno

- **Scopo del *Machine Learning*;**
- **Origini del *Machine Learning*;**
- **La *maledizione della dimensionalità*;**
- **Confidenza delle previsioni ottenute.**

3 Resoconto

La riunione, da considerarsi più un seminario, è stata indetta a seguito dell'avvio del secondo lotto. L'azienda proponente, rappresentata dal *Dott. Gregorio Piccoli*, ha esposto quindi la tematica del *Machine Learning* trattandone non solo gli aspetti inclusi nel capitolato, ma dandone una visione complessiva; sono stati inoltre forniti alcuni casi esempio che hanno reso il tutto maggiormente comprensibile.

Il gruppo *ProApes* non ha potuto partecipare in modo simultaneo all'incontro a causa di alcuni problemi tecnici; tuttavia essendo stata registrata i membri hanno comunque potuto visionare la riunione in un secondo momento, preservandosi di contattare il proponente nel caso di dubbi o incomprensioni.

Di seguito è riportata la lista degli argomenti affrontati.

3.1 Scopo del *Machine Learning*

Il *Dott. Gregorio Piccoli* ha ribadito il concetto di come il *Machine Learning* a partire da un certo insieme di dati osservati, fornisca grazie all'ausilio di appositi algoritmi, delle regole in grado poi di fare delle previsioni. La potenza di questo metodo sta proprio in questo, ovvero prendendo in esame un certo sottoinsieme della popolazione si definiscono regole valide per la sua totalità. Un esempio portato che ha fatto capire molto bene tale concetto riguarda gli esami del sangue: grazie all'estrazione di poche gocce, si ottengono risultati validi per l'intera fisiologia del paziente.

Per fare tutto ciò un ruolo fondamentale lo gioca l'addestramento del sistema; pensiamo a tal proposito al calcolo dell'indice di massa corporea, tale calcolo fa uso di un algoritmo di *Machine Learning* dove data l'altezza (e magari anche il sesso) della persona se ne è in grado di prevedere il peso. Attraverso quindi un qualcosa che si ottiene con molta facilità, l'altezza, se ne estrapola il peso grazie all'ausilio di un predittore anch'esso di costo ridotto; sarà sufficiente quindi fornire al sistema l'altezza per poter prevedere il peso esatto.

3.2 Origini del *Machine Learning*

Il tema di come questa disciplina sia nata e di quali siano le sue sfumature, sono state trattate seppur superficialmente in modo esaustivo per le circostanze, toccando anche temi non richiesti specificatamente dal capitolato.

Il *Machine Learning* deve le sue origini alla statistica e alla incapacità di aumentarne la precisione di calcolo anche a seguito dell'utilizzo di strumenti di potenza superiore. Chiamata *la maledizione della dimensionalità* tale problematica si deve al fatto che aumentando il numero di predittori, lo spazio che si ha a disposizione conseguentemente cresce facendo in modo che i punti risultino sempre troppo lontani far loro per permettere un calcolo più preciso; in risposta a ciò è nato quindi il *Machine Learning*. Esso si basa sul *criterio di minimizzazione del rischio* e comprende al suo interno tre grandi famiglie:

- ***Supervised Learning***: si fornisce a degli algoritmi l'indicazione di ciò che si vuole prevedere; è qui che si trovano le tecniche di *Regressione Lineare* ed *SVM* previste dal capitolato;
- ***Unsupervised Learning***: a differenza di sopra non si possiede più il target obiettivo; semplicemente vi sono dati e si cercano metodi che ne estraggano regole. Anche in tale categoria si trovano degli algoritmi quali la *Clusterizzazione*

che permette a partire da dei dati osservati di individuarne dei raggruppamenti precisi; e la *Riduzione di dimensionalità* dove è il sistema stesso che dice quali misure sono interessanti (elimina i predittori che non danno informazioni aggiuntive);

- **Reinforcement Learning**: permettono di predire o meno il verificarsi di un dato comportamento. I sistemi che rientrano in questa categoria imparano sbagliando a diventare sempre più bravi (e.g: giochi che sfruttano A.I).

3.3 La maledizione della dimensionalità

Sebbene questi temi inerenti *Regressione Lineare* ed *SVM* siano già stati affrontati dal proponente negli incontri precedenti, ora vengono discussi ponendo maggiore enfasi alla *maledizione della dimensionalità* e di come essa sia stata risolta dall'avvento dell'algoritmo di *SVM*.

La *Regressione Lineare* lavorando nel campo dei valori continui cerca la regola che minimizza la somma delle distanze fra i valori osservati; appare evidente come non riesca a spiegare tutto ma cerchi di trovare la retta passante il più vicino possibile alle misure rilevate. Per quanto concerne invece *SVM* consiste in algoritmi di classificazione che cercano attraverso un vettore di supporto, di trovare l'iperpiano in grado di dividere meglio la popolazione in classi differenti. L'efficienza di tale metodo sta nel fatto che non si lavora più in una realtà lineare ma con dei polinomi: più aumentano le dimensioni e più la tecnica *SVM* diventa precisa; vi è inoltre anche una certa elasticità nell'iperpiano in quanto farà tante curve quanti sono i gradi del polinomio stesso.

3.4 Confidenza delle previsioni ottenute

L'ultimo argomento trattato riguarda i problemi di *underfitting* e *overfitting* e di come riuscire a capire se la previsione ottenuta sia corretta o meno:

- **underfitting**: si verifica quando si applicano algoritmi troppo rigidi (con pochi parametri) rispetto ai dati osservati, come quando si cerca di applicare una regola lineare ad una realtà che non lo è;
- **overfitting**: qui gli algoritmi usati sono troppo deboli rispetto ai dati osservati (con troppi parametri). Conseguentemente tale algoritmo risulterà perfetto se impiegato nei dati di test, ma poco praticabile nella realtà.

Il *Dott. Gregorio Piccoli* ha poi evidenziato come la giusta via sia data dall'individuazione di una via di mezzo: una regola abbastanza rigida ma non eccessivamente, che possa prevedere anche delle situazioni di errore seppur basse, e che funzioni non solo durante l'addestramento ma anche nella pratica.

Nonostante queste accortezze si sottolinea come non esistano algoritmi che vadano bene per tutto in quanto ciascuno di essi prende in esame delle caratteristiche, ma ne potrebbe escludere altre che potrebbero risultare egualmente interessanti. Spetta dunque a colui che applica tali algoritmi capire cosa è effettivamente rilevante e cosa no al fine di ottenere una previsione che sia il più corretta possibile, in quanto le macchine non operano alcuna distinzione.

4 Riepilogo delle decisioni

Codice	Decisione
VE_2020-04-03.1	Esposizione dello scopo del <i>Machine Learning</i> .
VE_2020-04-03.2	Trattazione dei motivi che hanno portato alla nascita del <i>Machine Learning</i> e sua conseguente suddivisione in varie tipologie.
VE_2020-04-03.3	Trattazione della <i>maledizione della dimensionalità</i> con conseguente risoluzione grazie ad <i>SVM</i> .
VE_2020-04-03.4	Discussione circa i problemi di <i>underfitting</i> e <i>overfitting</i> per comprendere quale sia l'approccio di previsione maggiormente corretto da impiegare.

Tabella 2: Riepilogo delle decisioni