



Markers e collegamento tra terroir e prodotti caseari : metodologie e risultati

Giovanna Contarini CRA-FLC - Lodi;
Michele Forina DiCTFA - Università di
Genova

Matrici ed analisi

Analisi
bromatologiche

Rapporti
isotopici



3104

Costituenti
frazione
volatile

Costituenti
frazione
idrocarburica

Costituenti
oli
essenziali



Costituenti
di base

Acido
linoleico
coniugato

Comp.
microbica



6816

Rapporti
isotopici

Costituenti
frazione
volatile

Costituenti
frazione
idrocarburica



Costituenti
di base

Acido
linoleico
coniugato

Comp.
microbica

4224

Rapporti
isotopici

Parametri
reologici

Cost.
frazione
volatile



14144



Linea 2: individuare markers utili a stabilire le possibili relazioni tra prodotti caseari e base alimentare degli animali

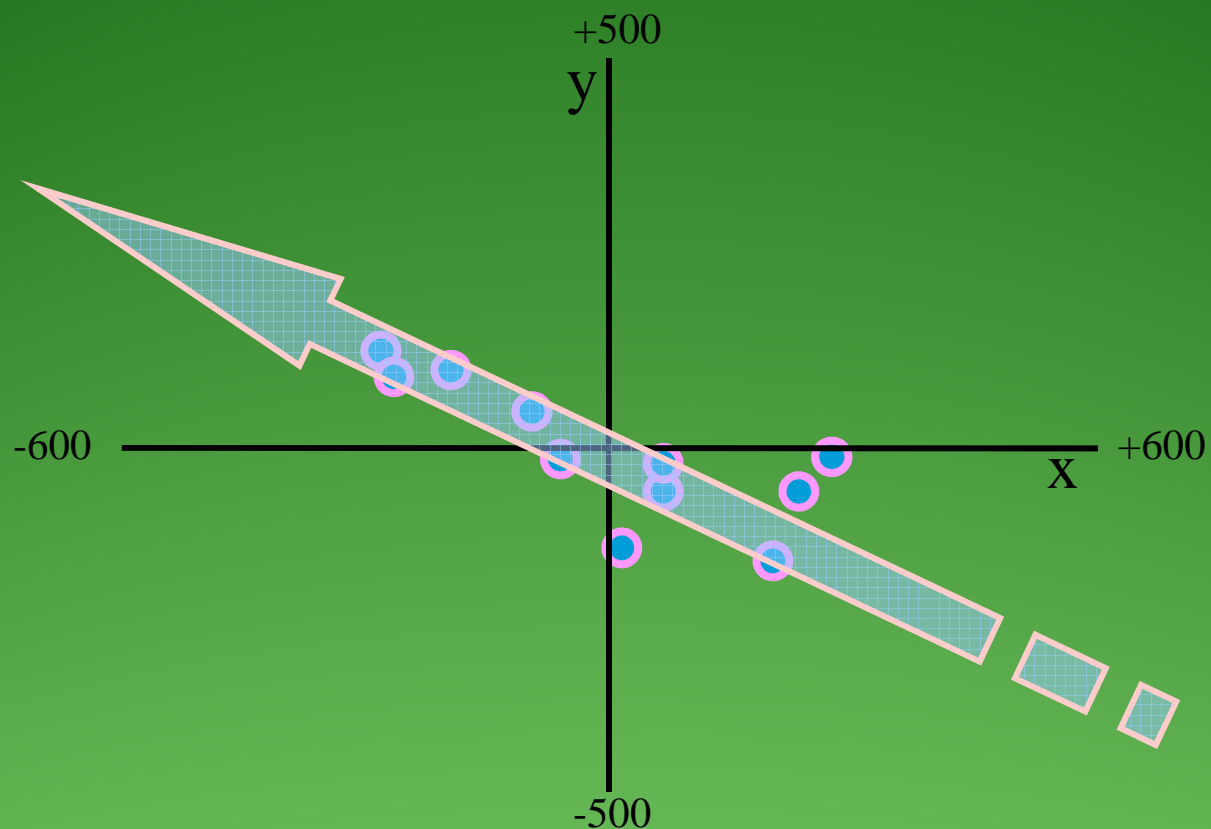
Applicare ai numeri ottenuti metodologie in grado di:

- **studiare i rapporti di correlazione tra le variabili**
- **visualizzare gli oggetti (campioni) tenendo conto di una larga frazione dell'informazione totale**
- **ridurre la dimensionalità dei dati, cioè descrivere i dati in termini di un numero minore di variabili rispetto a quelle originali che ne preservino tuttavia l'informazione sistematica e rilevante**
- **riconoscere campioni simili e/o differenti, quindi individuare la presenza di gruppi/categorie**
- **individuare dati anomali, cioè dati il cui comportamento si discosta nettamente da quello degli altri**

La PCA (analisi dei componenti principali) è il metodo più utilizzato in chemiometria per affrontare i problemi di analisi multivariata. Esso “*estrae*” l’informazione prevalente e sistematica delle variabili che descrivono il sistema.

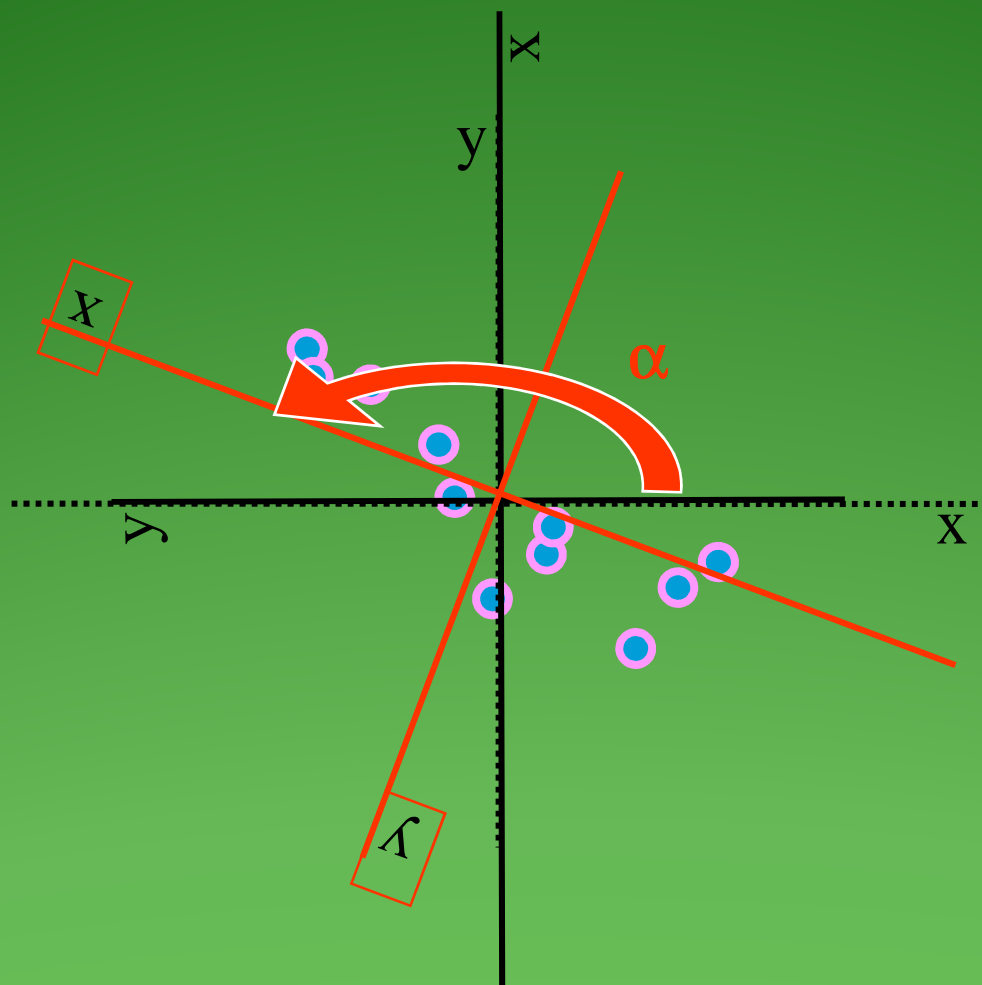
L’analisi dei componenti principali si basa sulla rotazione dello spazio originale in uno spazio costruito in modo che ciascuna nuova variabile (componente principale, PC) sia ortogonale rispetto alle altre e che ciascuna nuova variabile spieghi la massima frazione possibile di varianza dei dati non ancora spiegata dalle componenti principali precedenti.

L'esempio in un sistema bivariato



direzione di massima varianza cioè di
massima informazione

Rotazione degli assi in modo che l'asse della 1
variabile (x) corrisponda alla direzione di massima
varianza



Individuazione dell'angolo corrispondente alla rotazione effettuata

Grafico degli scores

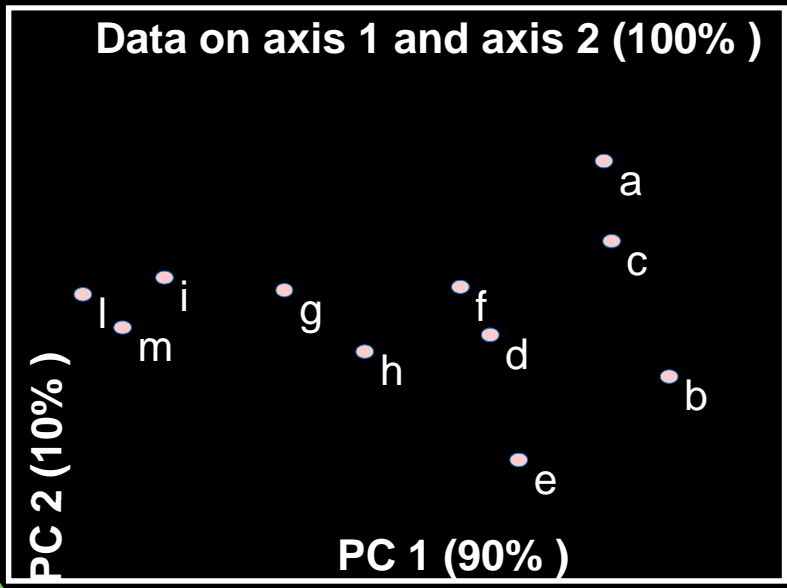
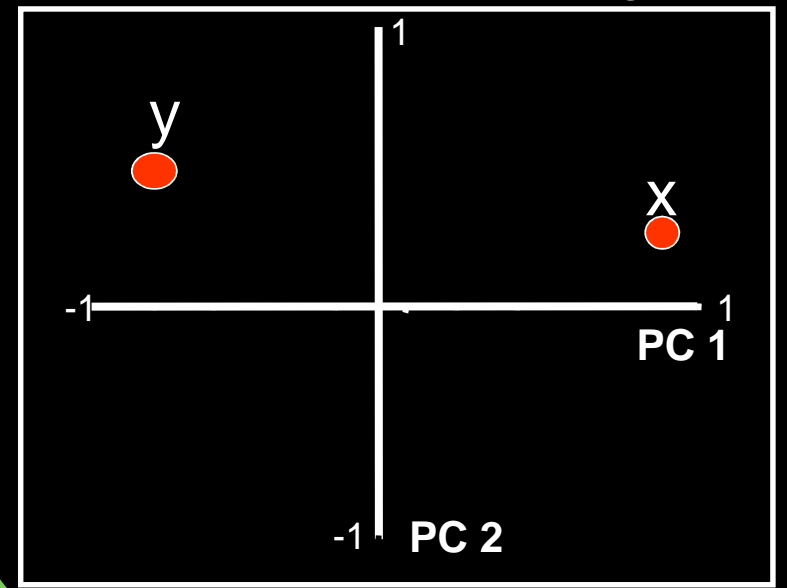
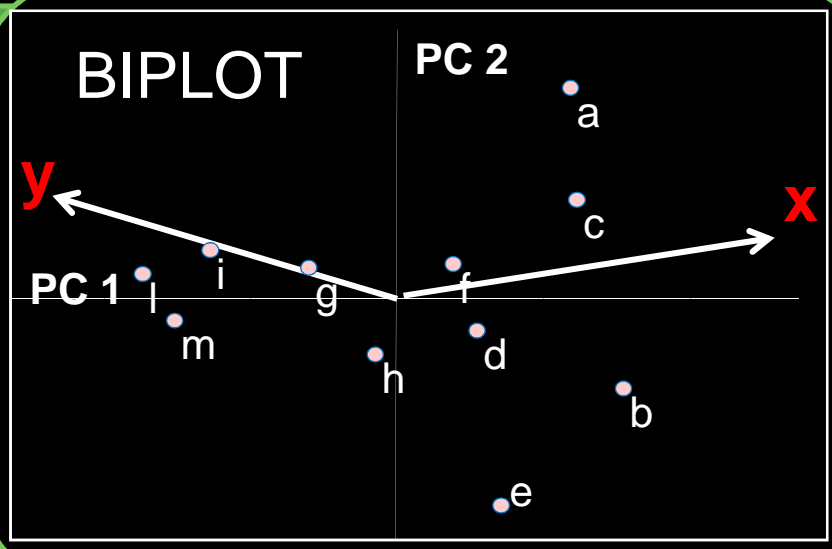


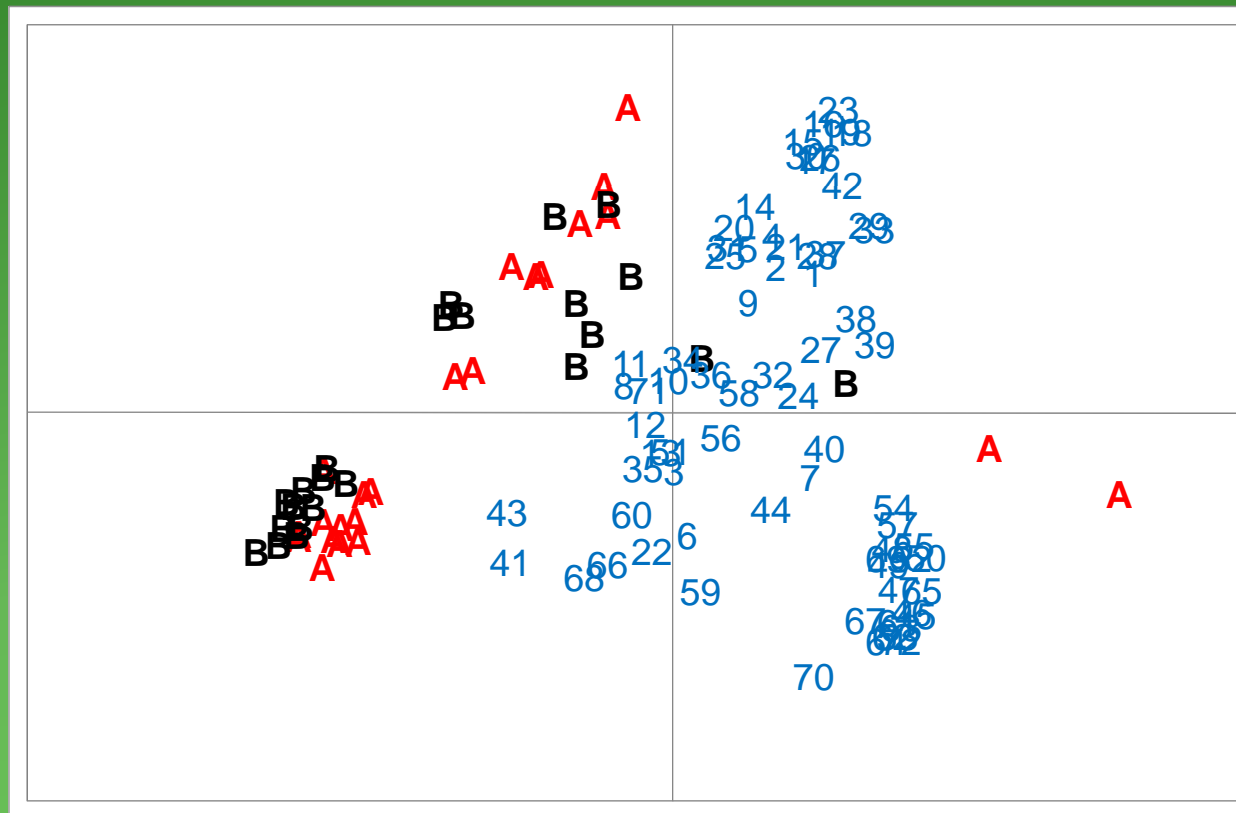
Grafico dei loadings



BIPLOT

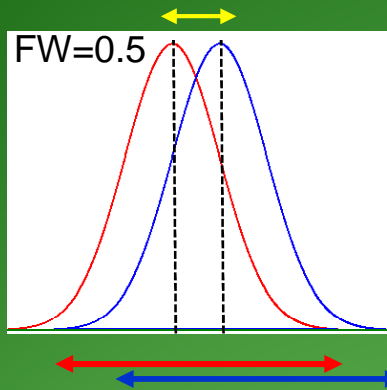


Esempio di PCA completa su una delle matrici:
 Formaggio: 48 campioni (A e B) e 72 variabili (48%
 varianza spiegata)



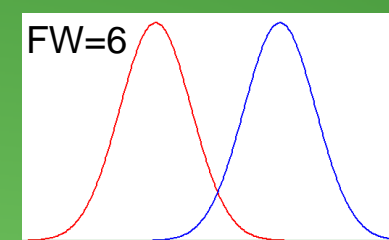
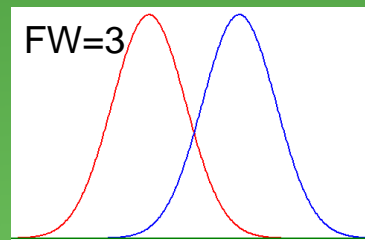
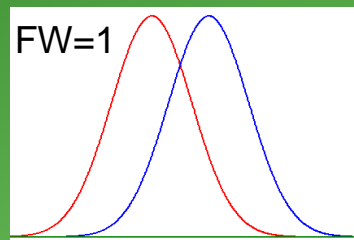
Selezione delle variabili

Per ogni variabile, per ogni tesi sperimentale (F e T, M e P) e per ogni anno, sono stati calcolati i Fisher Weights (FW), analoghi al rapporto di Fisher tra le varianze interclasse ed intraclasse.

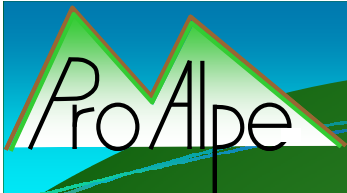


$$FW_v = \frac{(\bar{X}_{v1} - \bar{X}_{v2})^2}{\sum_{i=1}^{I_1} \frac{(X_{iv1} - \bar{X}_{v1})^2}{I_1} + \sum_{i=1}^{I_2} \frac{(X_{iv2} - \bar{X}_{v2})^2}{I_2}} =$$

■ Categoria 1 ■ Categoria 2



Il valore critico di FW è circa 0.38. Valori superiori indicano che la variabile è significativamente diversa nelle due categorie. Tuttavia valori di FW intorno a 1 corrispondono ancora ad una separazione piuttosto limitata

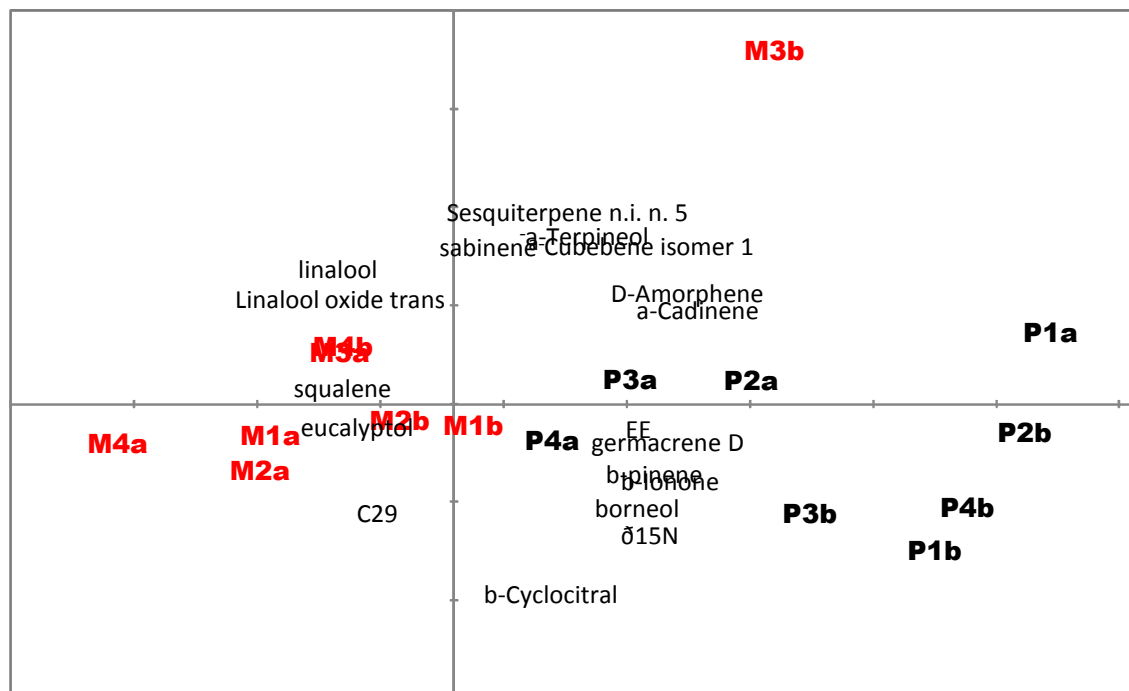


Valutazione sui campioni provenienti dal sito di Asiago

• Risultati relativi alle due facies M e P



PCA – 60% varianza spiegata con le prime due componenti



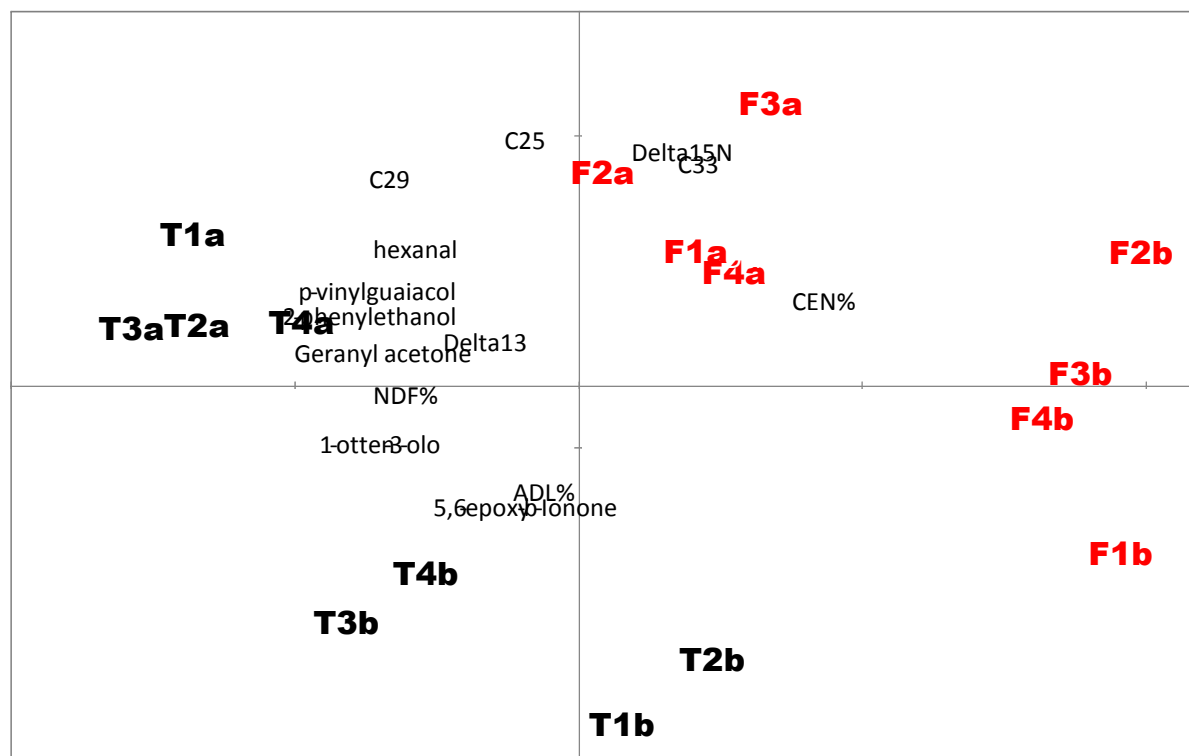
ASIAGO	Var.selezionate
Rapporti Isotopici	$\delta^{15}N$
Analisi di base	EE%
Volatili	α -Terpineol
Volatili	b-Ionone
Volatili	D-Amorphene,
Volatili	α -Cadinene
Volatili	α -Cubebene is. 1
Volatili	linalool oxide trans
Volatili	b-Cyclocitral
Volatili	sesquiter. n.i. n. 5
Oli essenziali	sabinene
Oli essenziali	b-pinene
Oli essenziali	Eucalyptol
Oli essenziali	Linalool
Oli essenziali	Borneol
Idrocarburi	C29
Idrocarburi	Squalene
Idrocarburi	germacrene D

Valutazione sui campioni provenienti dal sito di Valle Stura

• Risultati relativi alle due facies F e T



PCA – 58% varianza spiegata con le prime due componenti



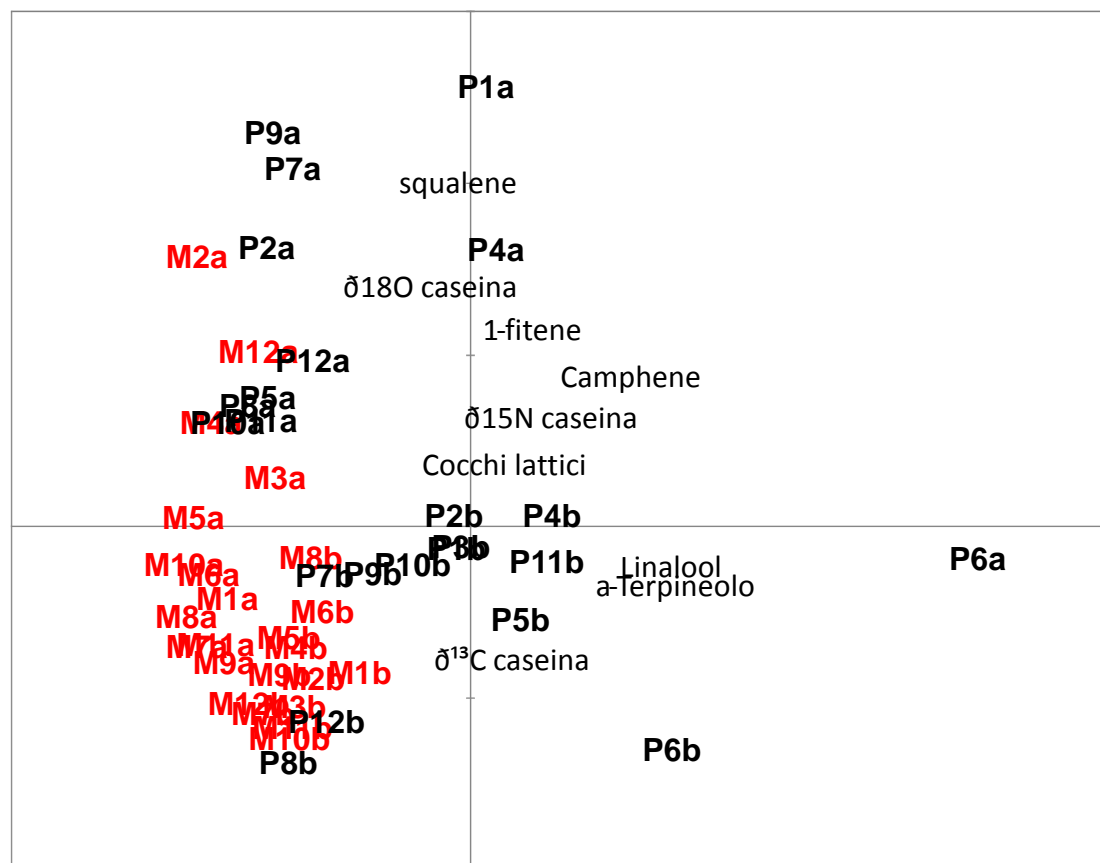
VALLE STURA	Var.selezionate
Rapporti Isotopici	$\delta^{15}\text{N}$
Rapporti Isotopici	$\delta^{13}\text{C}$
Rapporti Isotopici	δD
Analisi di base	CEN%
Analisi di base	NDF%
Analisi di base	ADL%
Volatili	Geranylacetone
Volatili	5,6epoxy-b-Ionone
Oli essenziali	2-phenylethanol
Oli essenziali	1-otten-3-olo
Oli essenziali	Hexanal
Idrocarburi	C25
Idrocarburi	C29
Idrocarburi	C33

Valutazione sui campioni provenienti dal sito di Asiago

• Risultati relativi ai campioni di formaggio derivanti da M e P



PCA – 48% varianza spiegata con le prime due componenti



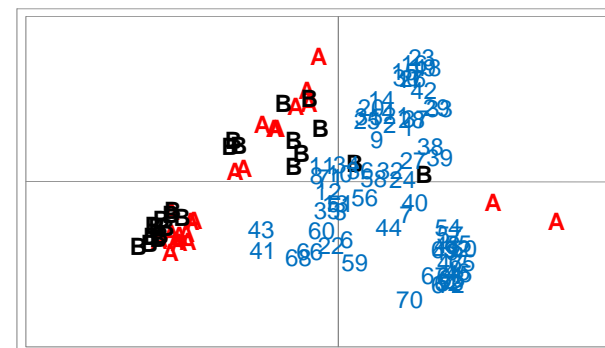
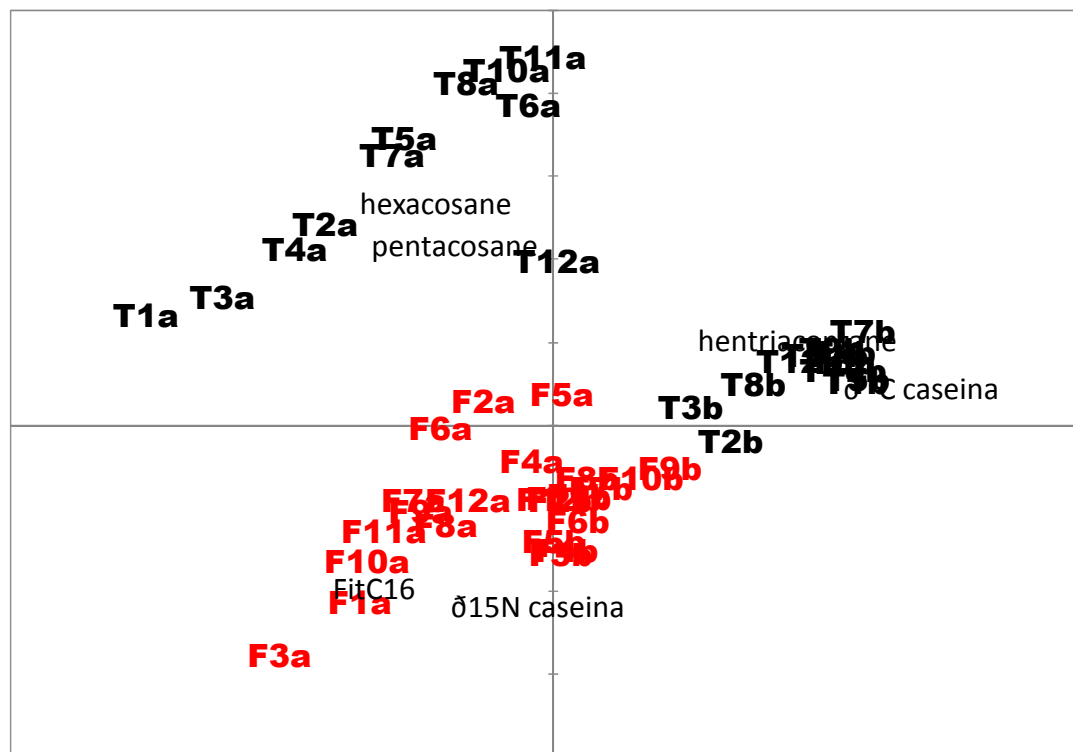
ASIAGO	Var.selezionate
Rapporti Isotopici	$\delta^{15}\text{N}$ caseina
Rapporti Isotopici	$\delta^{13}\text{C}$ caseina
Rapporti Isotopici	$\delta^{18}\text{O}$ caseina
Volatili	α -Terpineol
Volatili	Linalool
Volatili	Camphene
Idrocarburi	1-Fitene
Idrocarburi	Squalene
microbiol.	Cocchi lattici

Valutazione sui campioni provenienti dal sito di Valle Stura

- Risultati relativi ai campioni di formaggio derivanti da F e T



PCA – 66% varianza spiegata con le prime due componenti



VALLE STURA	Var.selezionate
Rapporti Isotopici	$\delta^{15}N$ caseina
Rapporti Isotopici	$\delta^{13}C$ caseina
Idrocarburi	C25
Idrocarburi	C26
Idrocarburi	C31
Idrocarburi	FitC16

Mediante PCA si può “*estrarre*” l’informazione contenuta nelle variabili che descrivono il sistema. La PCA è un’analisi ESPLORATIVA in cui i dati non sono necessariamente categorizzati quindi:

NON classifica

NON crea modelli

Ma è certamente la base per applicare tecniche di classificazione e modellamento

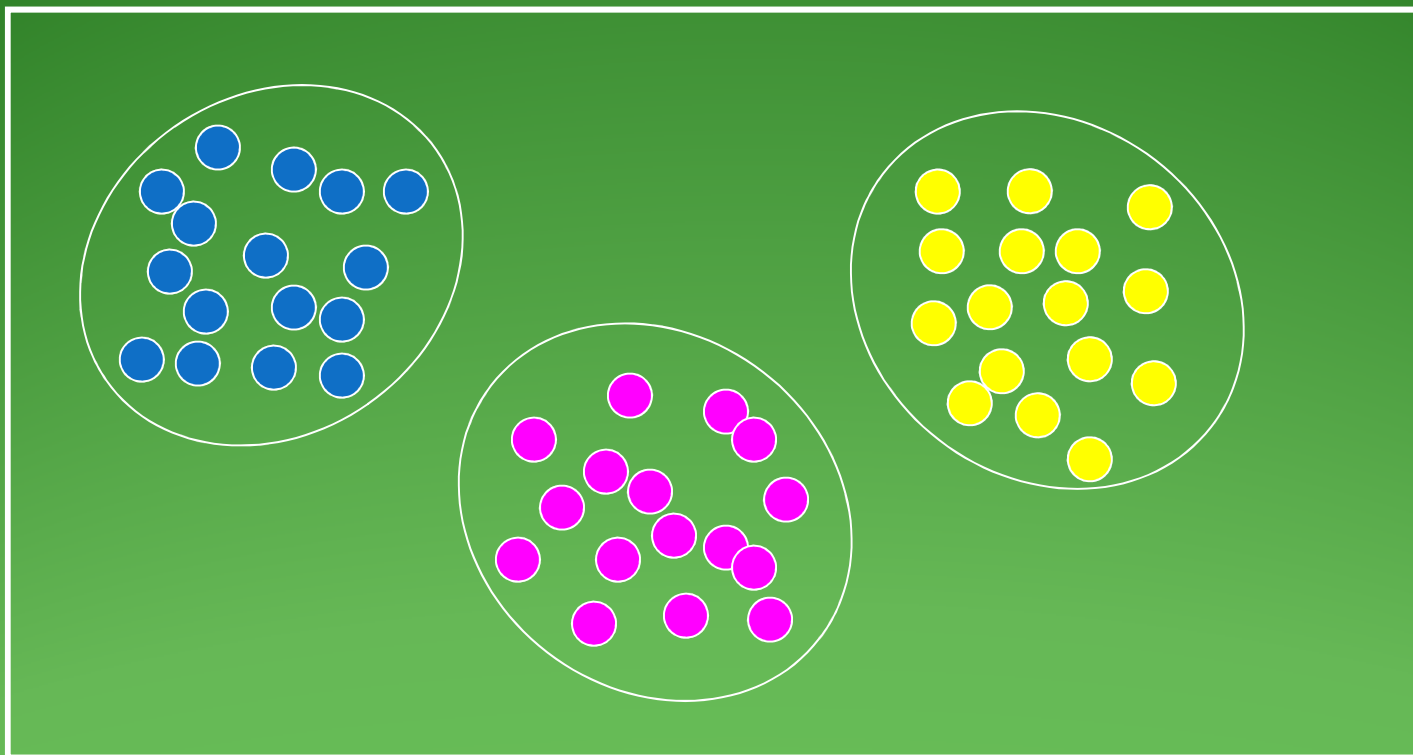
Al contrario i metodi di classificazione e modellamento necessitano di una conoscenza "A PRIORI" delle caratteristiche del dataset, in particolare l’esistenza di categorie e l’appartenenza di un determinato gruppo di oggetti ad una o ad un’altra categoria.

Obiettivo : individuazione di marker di territorio nel prodotto finale, cioè il formaggio



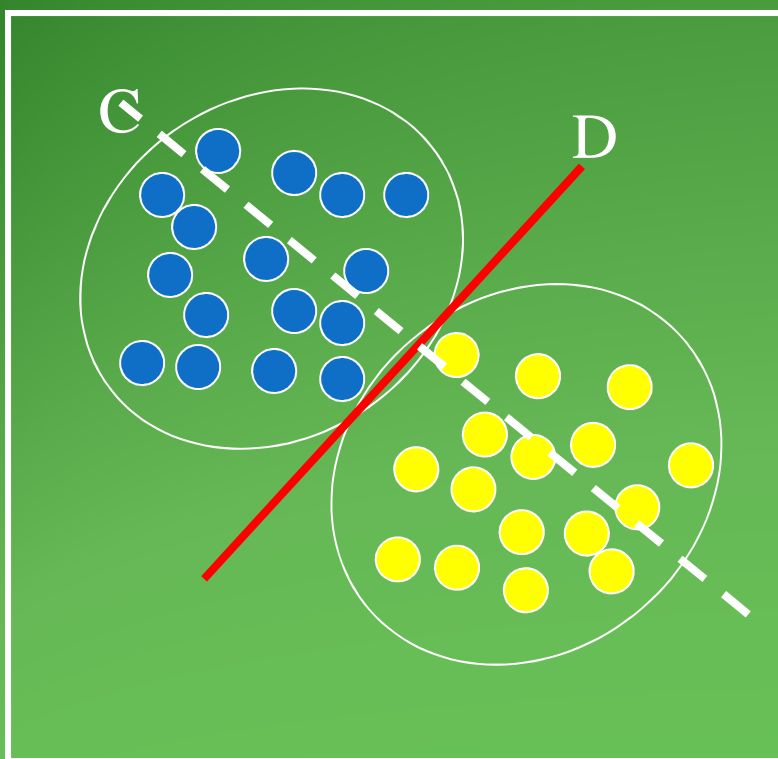
- **Ulteriori tecniche di selezione delle variabili:** SLDA (Stepwise LDA) o SELECT (metodo di decorrelazione progressiva)
- **tecniche di classificazione:** LDA (Linear Discriminant Analysis) e DPLS (Discriminant Partial Least Squares) QDA (Analisi discriminante quadratica)
- **idonee procedure di validazione:** convalida Leave-one-Out e convalida completa (CoVa) nel caso della selezione di variabili

Si basa sull' ipotesi che le distribuzioni all' interno delle categorie siano normali e le covarianze simili cioè le categorie hanno centroidi differenti ma dispersione simile intorno al centroide.



Centroide= origine delle coordinate ortogonali del baricentro cioè del centro di massa del sistema

LDA individua la direzione che massimizza il rapporto tra la covarianza inter-categoria e quella intra-categoria [C]. Questa è una direzione di massima discriminazione che descrive una funzione discriminante. La linea o il piano che interseca ortogonalmente questa direzione è definito come delimitatore [D] delle categorie



La proiezione degli oggetti sulla funzione discriminante permette di calcolare dei punteggi o voti discriminanti per ciascun oggetto

L' oggetto ignoto viene assegnato alla categorie per la quale presenta il minor punteggio discriminante.



Oggetti che vengono utilizzati per descrivere una categoria
TRAINING SET

ABILITA' DI CLASSIFICAZIONE: percentuale di oggetti del training set correttamente attribuiti alla categoria di appartenenza.

Oggetti che vengono utilizzati per validare le caratteristiche di una categoria :**EVALUATION SET** (categoria nota, ma inseriti come incogniti)

ABILITA' DI PREDIZIONE: percentuale di oggetti che non sono stati usati per costruire il modello, ma vengono correttamente classificati

Oggetti incogniti che devono essere assegnati ad una categoria
TEST SET



Valutazione sui campioni provenienti dal sito di Asiago

- Risultati relativi ai campioni di formaggio derivanti da M e P

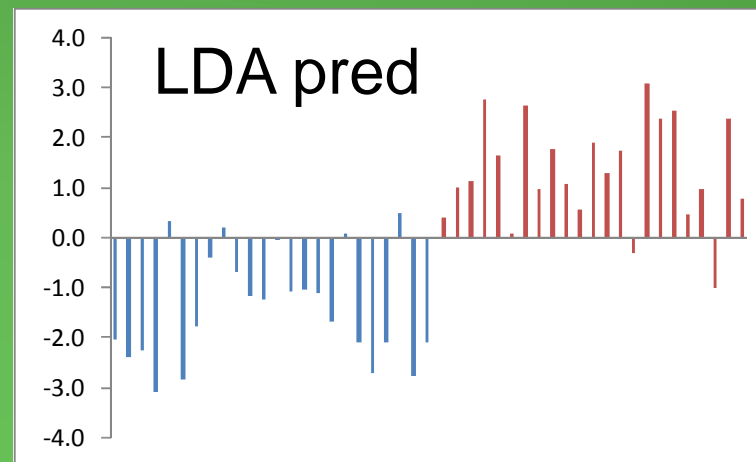
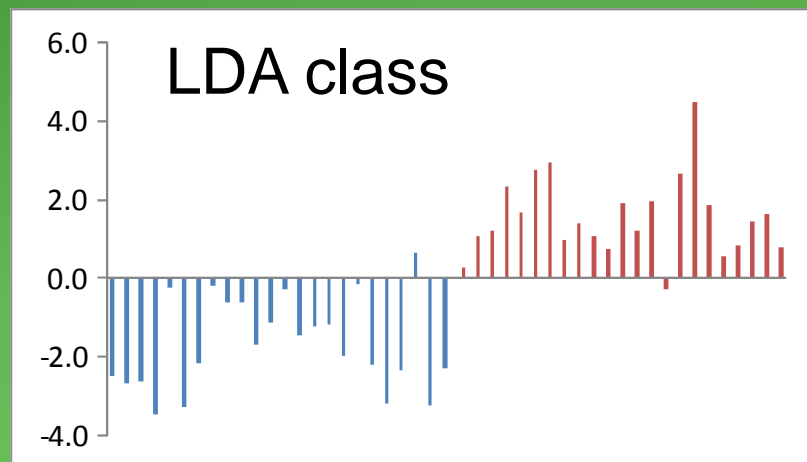


LDA con 10 variabili (micr., volatili, idroc, isotopi)

Risultati in classificazione: 2 oggetti non correttamente assegnati (4% errore)

Risultati in predizione: 6 oggetti non correttamente assegnati (13% errore)

 Magro  Pingue

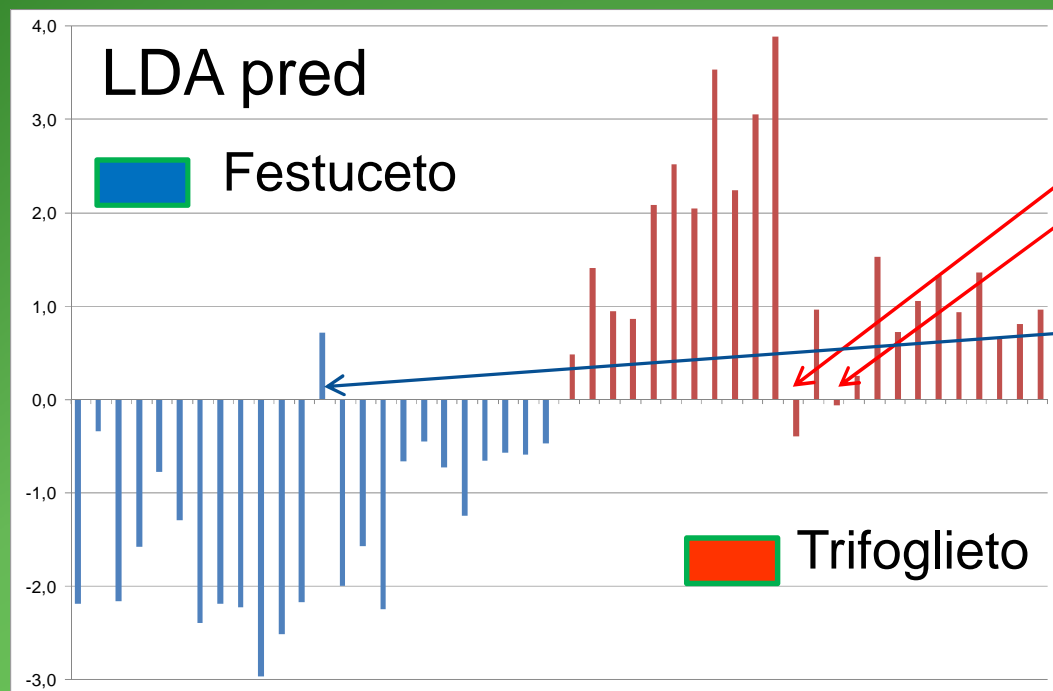


Valutazione sui campioni provenienti dal sito di Valle Stura

- Risultati relativi ai campioni di formaggio derivanti da F e T



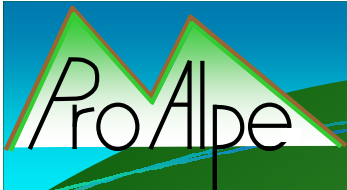
I migliori risultati in termini di abilità di predizione (6% di errore), con convalida completa si sono ottenuti tramite LDA utilizzando 4 costituenti della frazione idrocarburica: hexacosane, FitC16, hentriacontane, heptacosane



Tre campioni,
2 di Trifoglieto

1 di Festuceto

vengono assegnati alla categoria errata

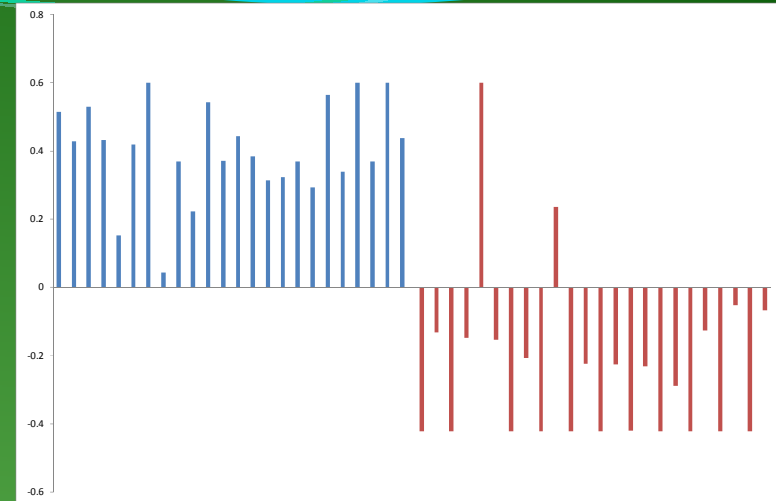


Variabili con capacità discriminante tra facies, ma spesso anche tra annate

VALLE STURA	Erba	Latte	Formaggio
Rapporti Isotopici	δ15N	δ15N caseina	δ15N caseina
Rapporti Isotopici	δ13C	δ13C caseina	δ13C caseina
Rapporti Isotopici	δD	δ18O acqua	
Rapporti Isotopici		δ18O caseina	
Analisi di base	CEN%		
Analisi di base	NDF%		
Analisi di base	ADL%		
Volatili	Geranylacetone		
Volatili	5,6epoxy-b-Ionone		
Oli essenziali	2-phenylethanol		
Oli essenziali	1-otten-3-olo		
Oli essenziali	Hexanal		
Idrocarburi	C25		C25
Idrocarburi	C29		C26
Idrocarburi	C33		C31
Idrocarburi			FitC16

ASIAGO	Erba	Latte	Formaggio
Rapporti Isotopici	δ15N	δ15N caseina	δ15N caseina
Rapporti Isotopici		δ18O acqua	δ13C caseina
Rapporti Isotopici		δD caseina	δ18O caseina
Analisi di base	EE%		Proteine
Volatili	a-Terpineol	a-Terpineol	a-Terpineol
Volatili	b-Ionone	b-Ionone	
Volatili	D-Amorphene,	b-caryophyllene	Methylsalicylate
Volatili	a-Cadinene	Cedrane	Linalool
Volatili	a-Cubebene is.	Limonene	b-caryophyllene
Volatili	linalool oxide trans	Linalool	a-Pinene
Volatili	b-Cyclocitral	(Z)-b-Ocimene	Camphene
Volatili	sesquiter.n.i. n. 5	p-Cymene	
Volatili		b-Pinene	b-Pinene
Volatili		Terpinolene	4-Terpineol
Volatili		Sabinene	b-Mircene
Volatili		Myrtenol	Myrtenol
Oli essenziali	sabinene		
Oli essenziali	b-pinene		
Oli essenziali	Eucalyptol		
Oli essenziali	Linalool		
Oli essenziali	Borneol		
Idrocarburi	C29		1-Fitene
Idrocarburi	Squalene		2-Fitene
Idrocarburi	germacrene D		isopropilC14
Idrocarburi			C17
Idrocarburi			Phytane
Idrocarburi			Squalene
micro			Cocchi lattici

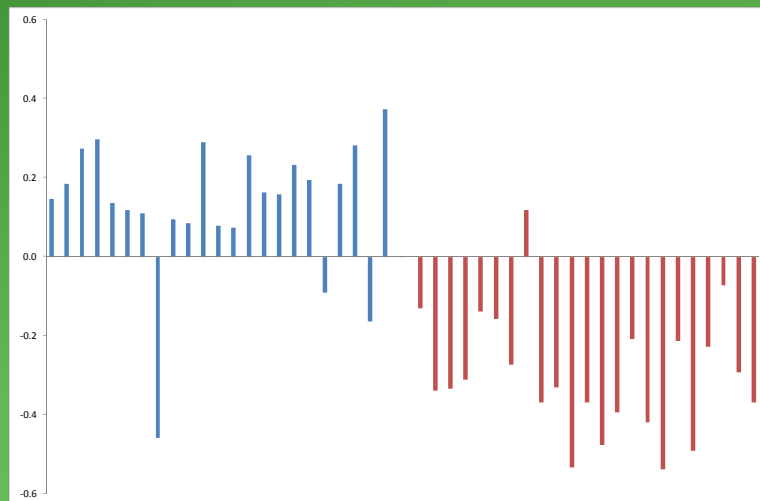
-Verifica dell'abilità discriminante di alcune trasformate (rapporti, differenze, prodotti) delle variabili originali per trovare parametri di semplice applicazione, ottenibili a costi ridotti e meno influenzati dall'anno di produzione



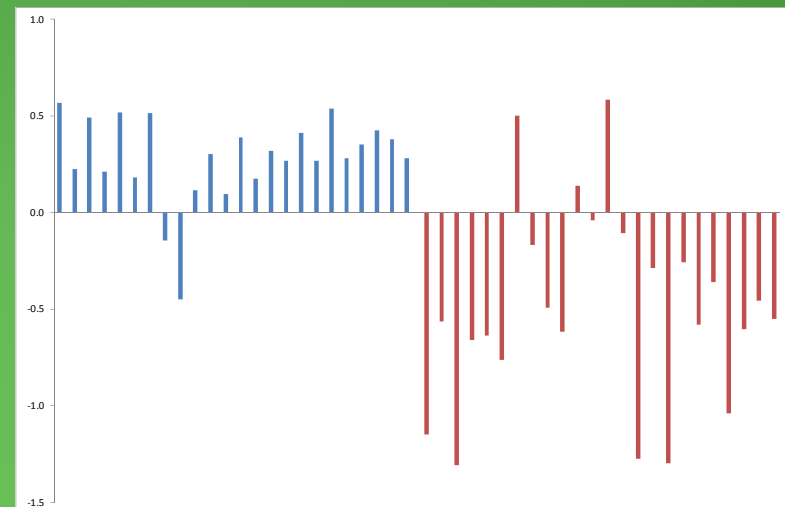
Trifoglieto Festuceto

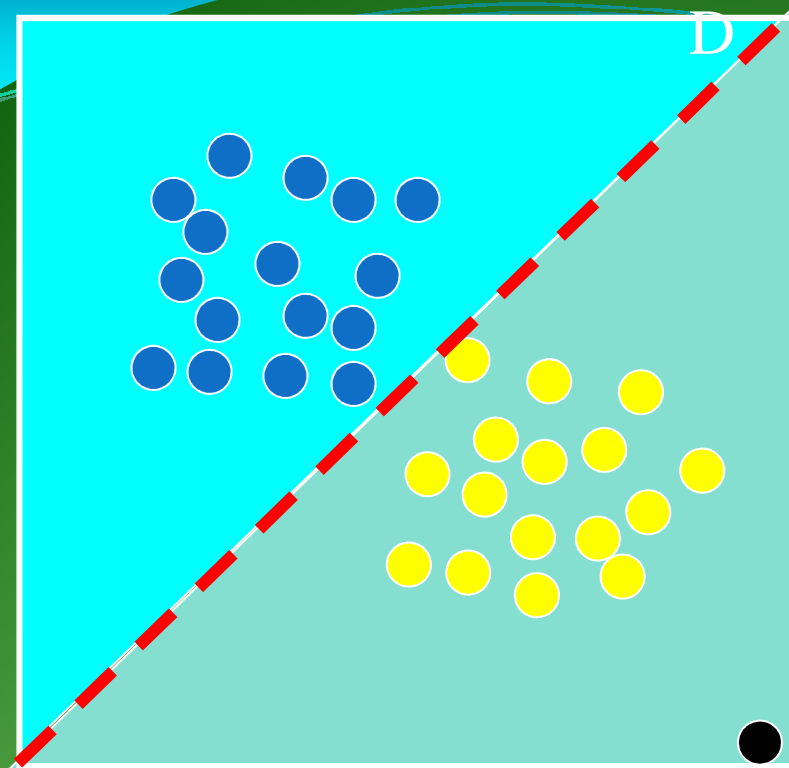
1-fitene/2-fitene

C29/C27



Magro Pingue

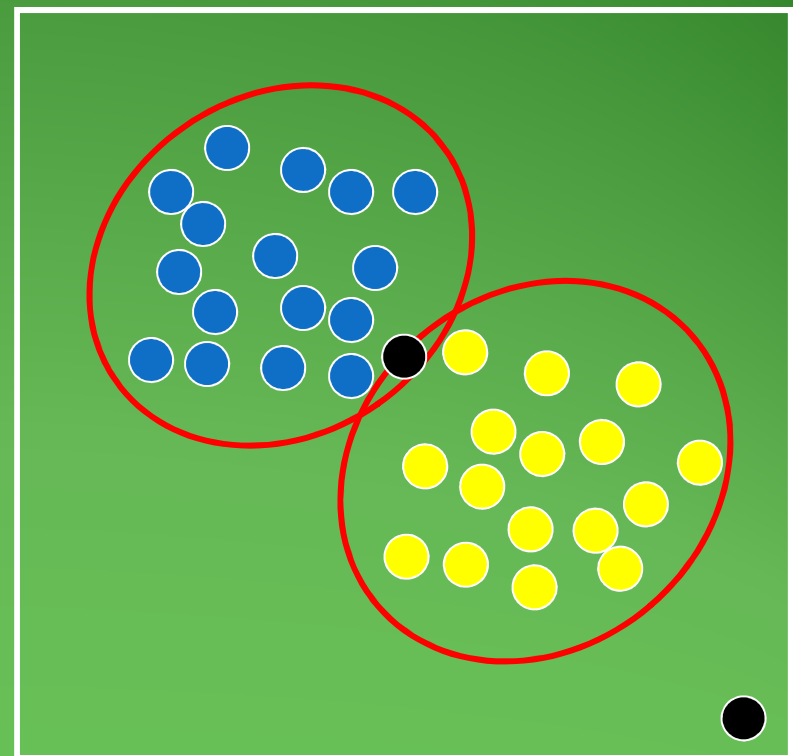




LDA: Primo livello della Pattern Recognition : solo il delimitter separa le due categorie –

Classificazione obbligata

Secondo livello: per ogni categoria è definito uno spazio; esiste quindi uno spazio non assegnato e può esistere uno spazio condiviso



Conclusioni

Markers



Collegamento tra terroir e prodotti caseari



Modello esportabile

