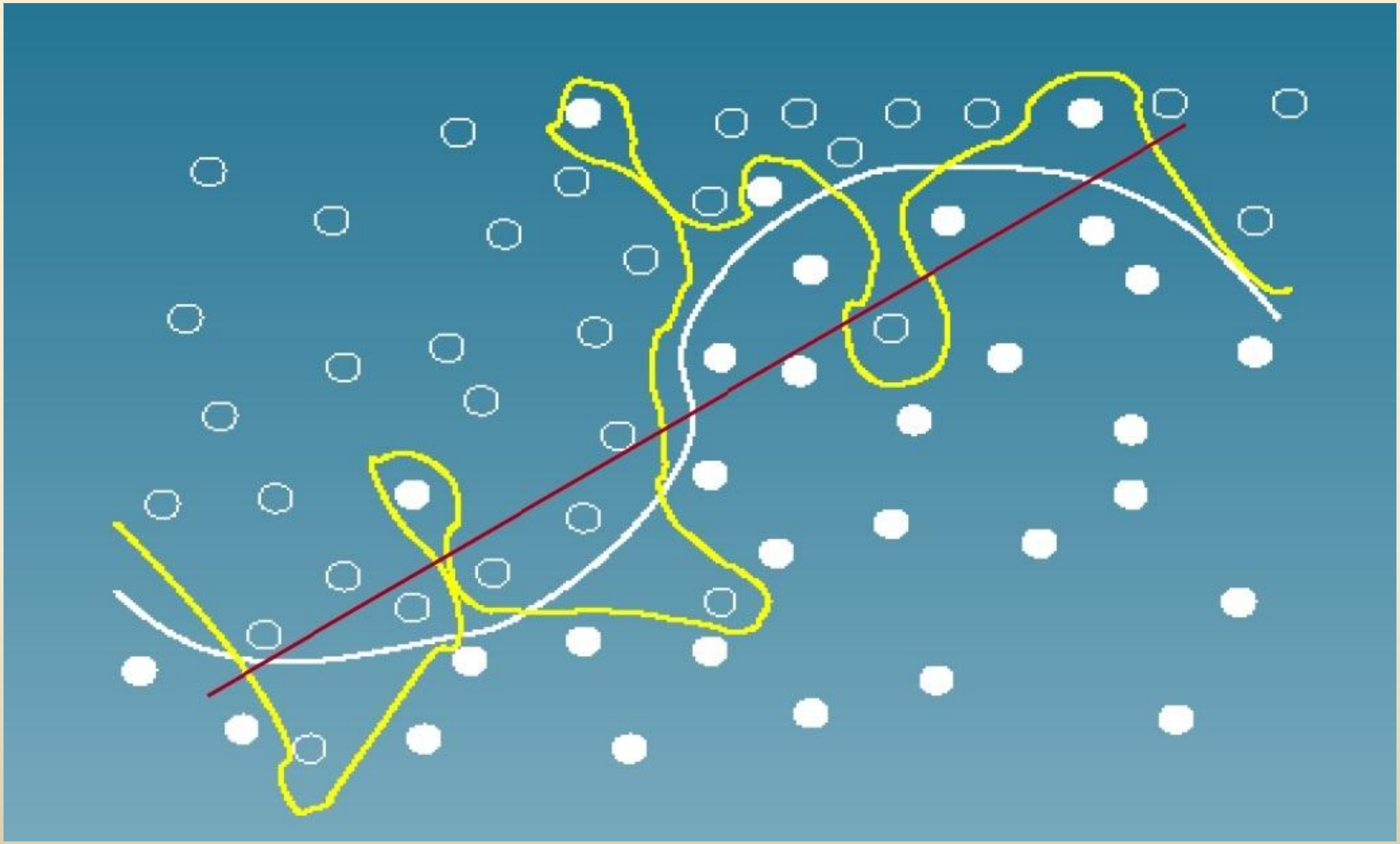
Laboratorio di Apprendimento Automatico

Fabio Aiolli
Università di Padova

Underfitting e Overfitting



Complessità spazio ipotesi

- SVM: aumenta con kernel non lineari, RBF con maggiore pendenza, aumentando parametro C (C bassi significa alto margine, bassa complessità)
- ALBERI DI DECISIONE: aumentando numero di livelli dell'albero
- KNN: aumenta al diminuire di K

Bias e Varianza

Il BIAS misura la *distorsione* di una stima
La VARIANZA misura la *dispersione* di una stima

$$b = \mathbb{E}[\hat{\theta}] - \theta$$
$$v = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$$

Model Selection and Hold-out

- Most of the time, the learner is parametric. These parameters should be optimized by testing which values of the parameters yield the best effectiveness.

- Hold-out procedure
  1. A small subset of Tr, called the validation set (or hold-out set), denoted Va, is identified
  2. A classifier is learnt using examples in Tr-Va.
  3. Step 2 is performed with different values of the parameters, and tested against the hold-out sample

- In an operational setting, after parameter optimization, one typically re-trains the classifier on the entire training corpus, in order to boost effectiveness (debatable step!)

- It is possible to show that the evaluation performed in Step 2 gives an unbiased estimate of the error performed by a classifier learnt with the same parameters and with training set of cardinality $|Tr|-|Va|<|Tr|$

K-fold Cross Validation

- An alternative approach to model selection (and evaluation) is the K-fold cross-validation method
- K-fold CV procedure
  - K different classifiers $h_1,h_2,\ldots,h_k$ are built by partitioning the initial corpus Tr into k disjoint sets $Va_1,\ldots,Va_k$ and then iteratively applying the Hold-out approach on the k-pairs $\langle Tr_i = Tr - Va_i, Va_i \rangle$
  - Effectiveness is obtained by individually computing the effectiveness of $h_1,\ldots,h_k$, and then averaging the individual results
- The special case $k=|Tr|$ of k-fold cross-validation is called leave-one-out cross-validation

Analisi Cross Validation

- Cosa succede al variare di K nella cross validation?
- K alto, training sets più grandi e quindi minore bias. Validation sets piccoli, quindi maggiore varianza.
- K basso, training sets più piccoli e quindi maggiore bias. Validation sets più grandi, quindi bassa varianza.

Titanic e classificazione NN

- Nell’esempio del TITANIC abbiamo utilizzato un classificatore Neirest Neighbor.
- Non ci siamo preoccupati di fare model-selection.
- Perchè?
  - Non avevamo parametri esterni.
  - Esempio di stima LOO dell’accuracy reale!

Evaluation for unbalanced data

- Classification accuracy:
  - usual in ML,
  - the proportion of correct decisions,
  - not appropriate if the population rate of the class is low
- Precision, Recall and F_1
  - Better measures

Contingency Table

	Relevant	Not Relevant
Retrieved	True positives (**tp**)	False positives (**fp**)
Not Retrieved	False negatives (**fn**)	True negatives (**tn**)

$$\pi = \frac{tp}{tp + fp} \qquad \rho = \frac{tp}{tp + fn}$$

Why NOT using the accuracy $\alpha = \dfrac{tp + tn}{tp + fp + tn + fn}$?

Effectiveness for Binary Retrieval: Precision and Recall

If relevance is assumed to be binary-valued, effectiveness is typically measured as a combination of

- Precision: the "degree of soundness" of the system
  - P(RELEVANT|RETURNED)
- Recall: the "degree of completeness" of the system
  - P(RETURNED|RELEVANT)

F measure

- A measure that trades-off precision versus recall?
  F-measure (weighted harmonic mean of the precision and recall)

$$F_\beta = \frac{(1+\beta^2)\pi\rho}{\beta^2\pi+\rho}$$

$\beta < 1$ emphasizes precision!

$$F_1 = 2\frac{\pi\rho}{\pi+\rho}$$