Identifying dominant models when the noise context is known

C. Ferri, J. Hernández-Orallo, A. Martínez-Usó and M.J. Ramírez-Quintana

DSIC, UPV, València, Spain
{cferri, jorallo, admarus, mramirez}@dsic.upv.es

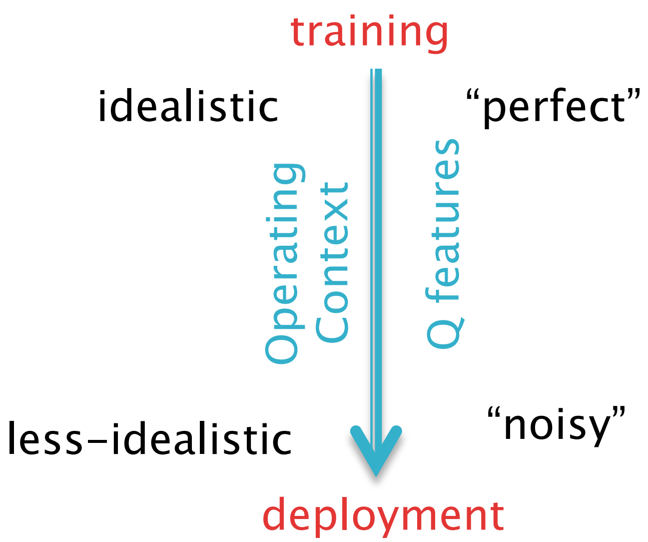
First International Workshop on Learning over Multiple Contexts
LMCE 2014
Nancy, 19 September 2014

Outline

- Motivation
- The Noise setting
- Context Plots and Dominance
- Experiments
- Conclusions and Future Work

Motivation

- Very often, operating contexts (OC) at the training and the deployment time are different.



- The error of the model depends on the level of noise introduced by the OC.

Motivation: An example

- Alarm system where two models (A and B) have been trained (ideal conds., OC=temperature in [0,30])
  - Validation: A>B (A is better!)
- Deployment
  - How these OC affect to sensors is also known
  - OC in deployment are given
- Which model is better for each OC? …. For several OC, model B could be better now!

Motivation

In order to answer this question we propose:

- To evaluate the models with different levels of simulate noise.
- To draw a context plot with all models, and to determine dominance regions.

- During deployment, the noise level is derived from the OC and the best model for that noise is applied.

The Noise setting

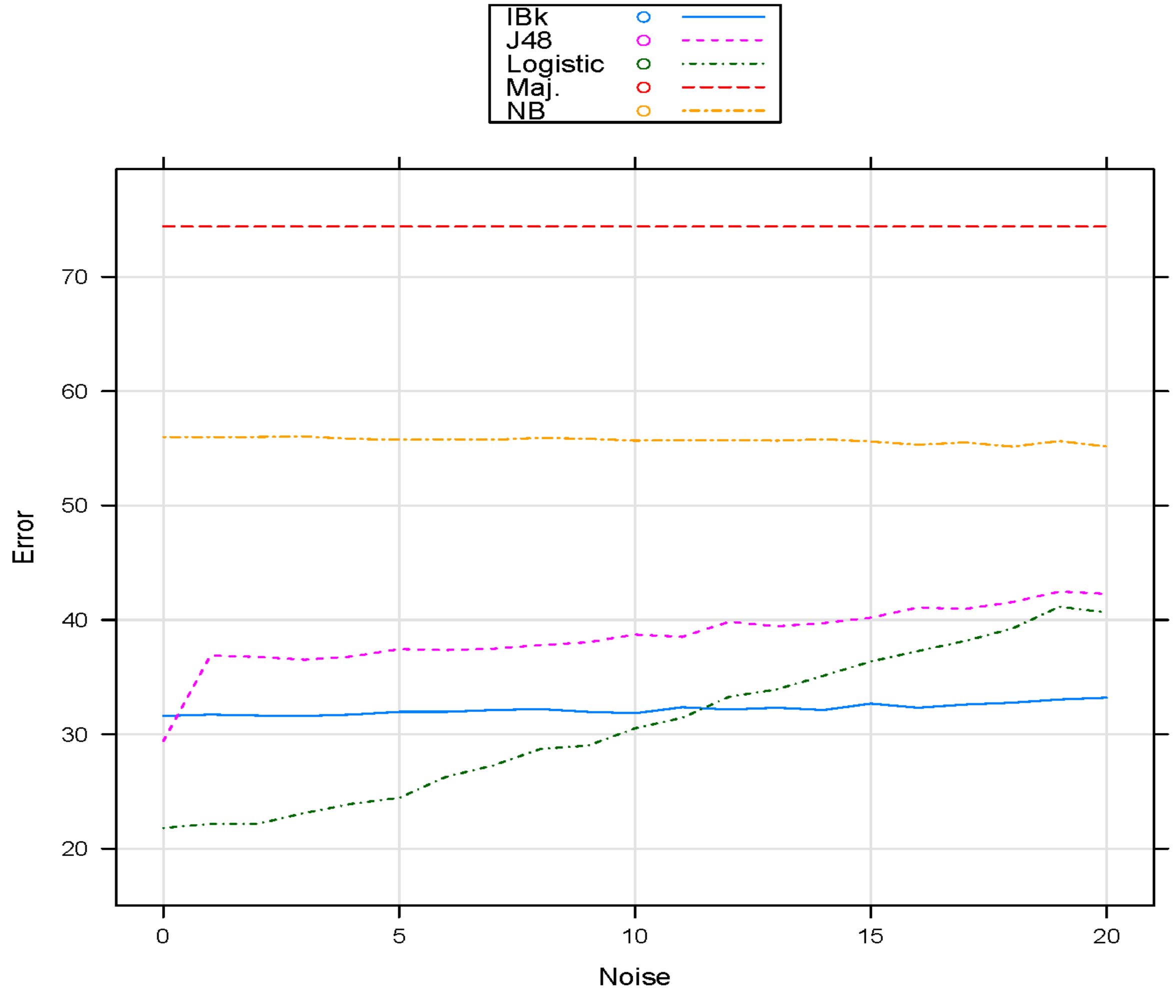
Noise is calculated by using probability distributions:

- Numerical attributes
  - we estimate the σ of all values of the attribute
  - for a level of noise ν, we modify a value x using a normal distribution $x' \sim N(x, \sigma \cdot \nu)$
- Nominal attribute
  - we estimate the frecuency of each value v_i, $p=(p_{v1},\ldots,p_{vn})$
  - for an instance x with value v_i we estimate the vector

    $$t=(t_1,\ldots,t_n) \quad t_i=1 \text{ and } t_j=0 \text{ if } i \neq j$$
  - for a level of noise ν we calculate $p' = \alpha \cdot p + (1-\alpha) \cdot t$ where $\alpha = 1 - e^{(-\nu)}$
  - we use p' to sample the new value x'

Context Plots and Dominance

Vehicle



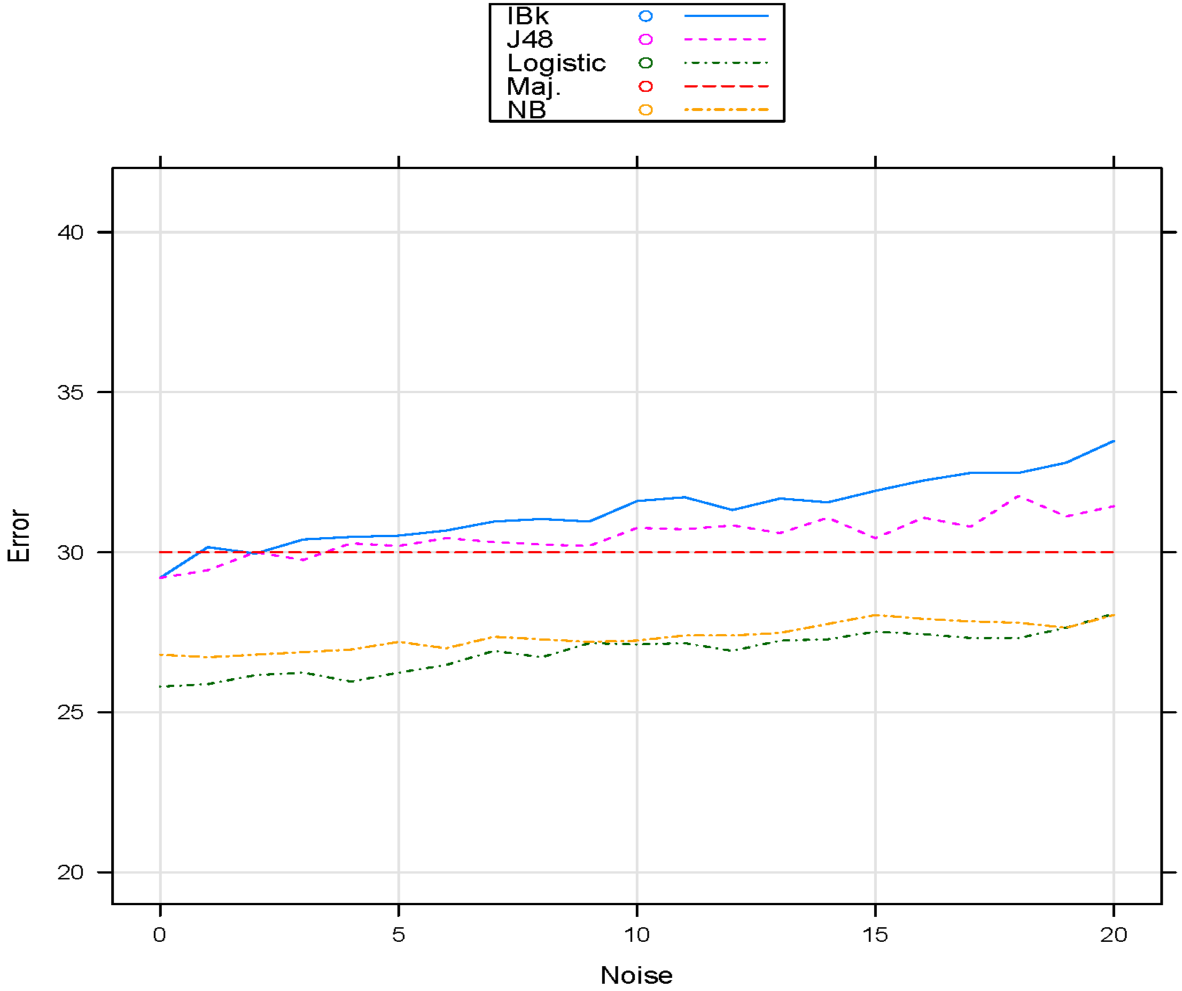
Experiments

Methodology:

- 12 datasets from the UCI repository
- 50% Train, 25% Validation, 25 % Test
- Classification:
    - J48, Naive Bayes, Logistic Regression and kNN.
    - Reference method: majority class.
    - Classification error
- Regression:
    - Linear Regression,M5P, kNN , SMOreg.
    - Reference method: ZeroR.
    - Relative absolute error

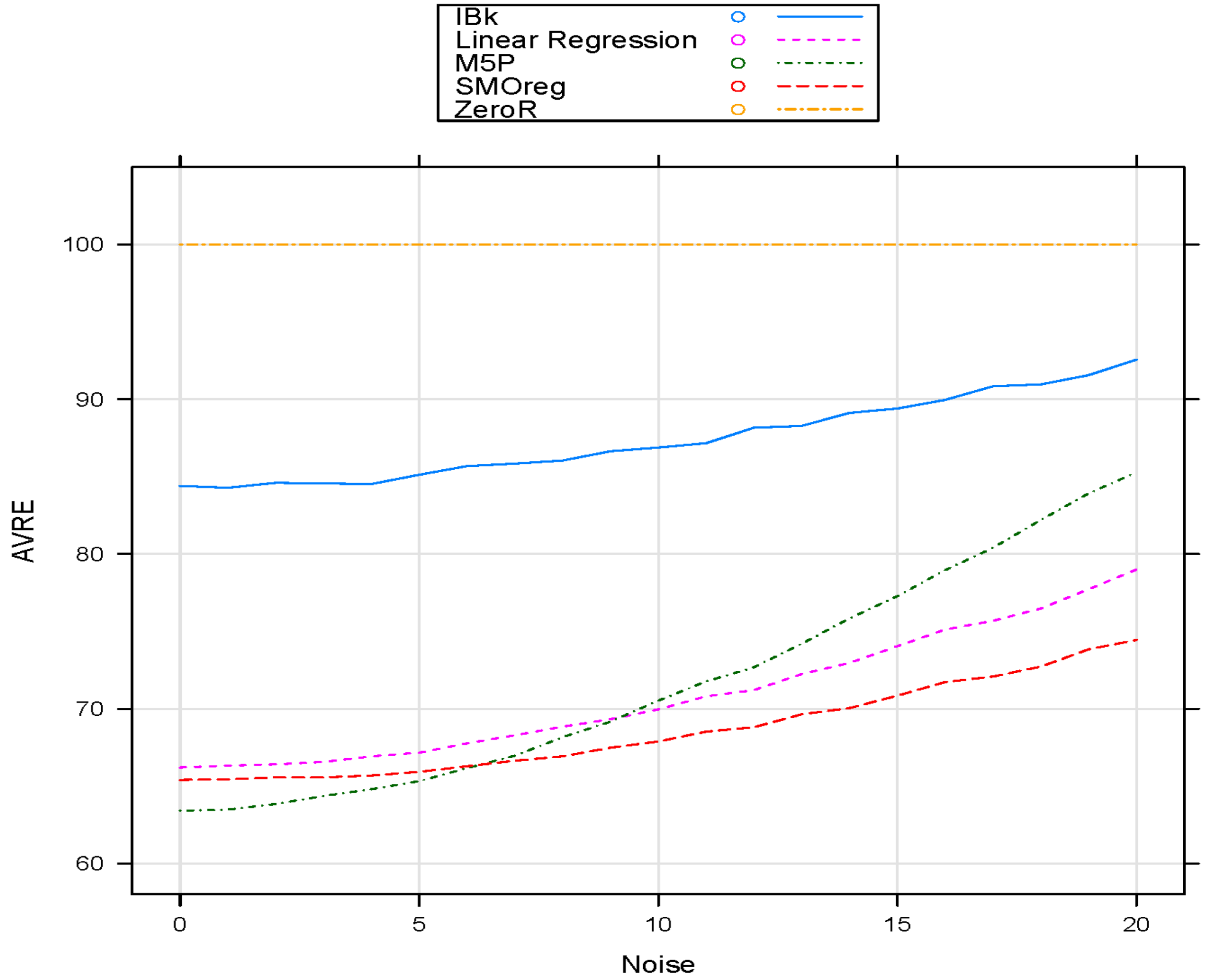
Context Plots and Dominance

Creditg



Context Plots and Dominance

Abalone



Experiments

Methods:

- **ValNoNoise**: For all the estimated values of noise, we select the method that obtains the best performance without noise.
- **ValBestArea**: For all the estimated values of noise, we select the method that obtains the best performance in the validation dataset by averaging all noise levels (i.e., the curve with lowest area in the context plot).
- **ValNoiseOpt**: For each of the estimated values of noise, we select the method with best performance in the validation dataset with that value of noise.
- **Idealistic**: For each of the estimated values of noise, we select the model with best performance in the test dataset with that value of noise. This strategy is not realistic (it cannot be done in practice). We just include this as a reference

Experiments

Dataset	ValNoNoise	ValBestArea	ValNoiseOpt
1 cholesterol	97.5468	97.5468	97.5468
2 bodyfat	99.2007	99.2007	99.9875
3 wisconsin	91.9391	91.9391	92.1545
4 stock	100.0000	100.0000	100.0000
5 meta	100.0000	100.0000	100.0000
6 abalone	81.6546	99.9358	99.9919
7 iris	89.3179	89.3179	96.4559
8 white-clover	89.9206	89.9206	88.3807
9 hepatitis	100.0000	100.0000	100.0000
10 breast-cancer	87.6362	87.6362	88.2716
11 vehicle	75.4220	96.1937	99.6372
12 credit-g	100.0000	100.0000	100.0000

- The performance results are normalised by the Idealistic performance

(idealistic/method)*100

Conclusions and Future Work

1. In this paper we have analysed the case when:

    - the noise level depends on a context
    - we know the context in advance

2. The model that best behaves for each noise level situation is used.

3. CONS: It takes some time. PROS: Selection/application of the best model is straightforward with results close to an idealistic process.

4. As a future work we plan working on different noise models, derived from scenarios with real operating conditions and with real sensors.

    - Each attribute will have a different operating range, but the context would still be given by a single parameter (e.g., temperature, or the number of measurements performed..).

Thanks for your atention…..