Computational Statistics
Lab II - linear regression

Luca Bortolussi

Department of Mathematics and Geosciences
University of Trieste

Office 238, third floor, H2bis
luca@dmi.units.it

Trieste, Winter Semester 2015/2016

Outline

1. Maximum Likelihood regression
2. Bayesian Regression

TASK 1

$E(w) = E_D(w) + \lambda \underbrace{w^T w}_{\|w\|^2}$

1) $RIDGE\_R(X, t, M, \lambda)$ Return w_{RR}
2) $RMSE(w, X_E, t_E)$
3) $CV(n, X, t, \lambda)$ return $RMSE(\lambda)$

FOR $\lambda \in$ ...
$e(\lambda) = CV(4, \ldots)$
PLOT(λ, e)

$w_{MLR} = (\lambda I + \Phi^T \Phi)^{-1} \Phi^T t$

$\log \lambda \in$

$\boxed{-20:0.1:5}$

- Consider the non-linear data of lab 1. Implement a regularised linear regression (ridge regression) with a polynomial model of degree $M = 12$.
- Set λ by n-fold cross-validation (choose a reasonable n), and by using a validation dataset. Compare the two approaches.

120

$n = 6$

20, 20, 20, 20, 20, 20

D_n

$RMSE(\lambda, D) = \frac{1}{n} \sum RMSE(\lambda, D \setminus D_n)$

do it for many λ, minimise RMSE.

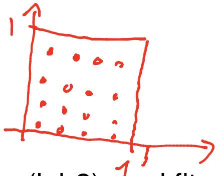
IN MATLAB. $\boxed{J = 30:60;}$

```
X1 = X;     X1(J) = [];     (X1 ∈ X \ X(J))
t1 = t;     t1(J) = [];

(OR   X1(J,:) = [];   FOR A MATRIX)

~> Xe = X(J);
~> te = t(J);
```

TASK 2



- Get the 2-dimensional data from moodle (lab2), and fit a model using gaussian basis functions:

$$\phi_i(\mathbf{x}) = \exp(-||\mathbf{x} - \mathbf{x_i}||^2/\gamma^2)$$

$\gamma \in [10^{-3}, 1]$

  Consider basis functions centered in a $k \times k$ regular grid in $[0, 1]^2$ (choose e.g. $k = 5$).
- Fit the model using ML, and use a validation set to identify the best γ.
- Fit the model using ridge regression, identifying both γ and λ by cross validation.