Linear and Logistic Regression

Machine Learning
uczeniemaszynowe@cs.uni.wroc.pl

28.10.2025, 10:00
exercise session

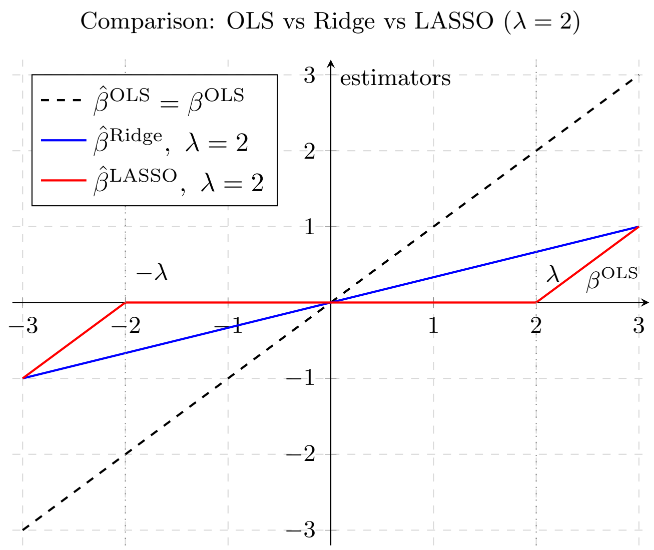

Figure 1: The relationship between regression coefficients for the method minimizing mean squared error, Ridge regularization, and LASSO.

Task 1 [2.5 pts]

Consider the dataset $\mathcal{D} = \{(y_i, x_i^{(1)}, x_i^{(2)})\}_{i=1}^{4}$, where

i	$x_i^{(1)}$	$x_i^{(2)}$	y_i
1	1	0	2
2	0	1	2
3	1	1	3
4	2	0	4

The linear regression model takes the form:

$$y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \varepsilon.$$

1.5 pts Compute the model coefficients:

- 0.5 pts by minimizing mean squared error,
- 0.5 pts with Ridge regularization,
- 0.5 pts with LASSO regularization.

1 pt Describe how regularization methods affect model parameters. Explain using Figure 1. Which regularization can be interpreted as a variable selection method?

Hints:

- Be ready to present your calculations on a different dataset during the class.
- Discuss smooth shrinking of regression coefficients toward zero and soft-thresholding.

Task 2 [3 pts]

The linear regression model takes the form:

$$y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \sigma)$.
For the dataset from Task 1, compute the maximum likelihood estimator (MLE) for:

- regression coefficients β,
- noise standard deviation σ or variance.

Solving one subtask gives 2 points; solving both yields 3 points total.
Hints:

- Recall the probability density function of the normal distribution and the definition of likelihood.
- Note that maximizing the likelihood is equivalent to maximizing its logarithm.
- Compute derivatives and find stationary points.
- Don't forget about the second derivative.
- The estimator $\hat{\beta}$ depends on y, and thus on ε, so it is noisy. The standard deviation of parameters $\sigma^{\hat{\beta}}$ is a linear transformation of the noise standard deviation σ (for fixed data).
- Be ready to compute on different data.

Task 3 [1.5 pts]

How can we check whether linear regression coefficients are statistically significant, i.e., whether a given explanatory variable has a meaningful effect on the dependent variable? Read the explanation below and test the significance of coefficients in the given example.

Introduction to Hypothesis Testing

Hypothesis testing is a statistical procedure that allows us to evaluate two competing statements about population parameters based on sample data.

Key Elements

- **Null Hypothesis (H_0):** assumes no effect or relationship, e.g. $\beta_j = 0$.
- **Alternative Hypothesis (H_1):** assumes a relationship exists, e.g. $\beta_j \neq 0$.
- **Significance Level (α):** commonly set to 0.05. It is the maximum allowable probability of a Type I error (rejecting a true H_0). We construct tests so as to minimize the probability of committing a Type II error (accepting a false null hypothesis H_0, i.e., failing to detect an existing effect) for a fixed significance level.
- **p-value:** probability of observing data as extreme or more extreme assuming H_0 is true.

Decision: If p-value $< \alpha$, reject H_0 and conclude statistical significance. Equivalently, we can define the set of extreme values of the test statistic for which we reject the null hypothesis.

Student's t-distribution

It is a continuous distribution used for inference when sample size is small and population variance is unknown (which is typical in regression).

- **Shape**: Symmetric and bell-shaped with heavier tails (higher probability of extreme values) than the normal distribution.
- **Degrees of freedom ν**: The shape of the distribution depends on a single parameter — the degrees of freedom. In linear regression, $\nu = n - p - 1$, where n is the number of observations and p is the number of predictors.
- **Asymptotic convergence**: As the number of degrees of freedom approaches infinity ($\nu \to \infty$), the Student's t-distribution becomes practically indistinguishable from the standard normal distribution.

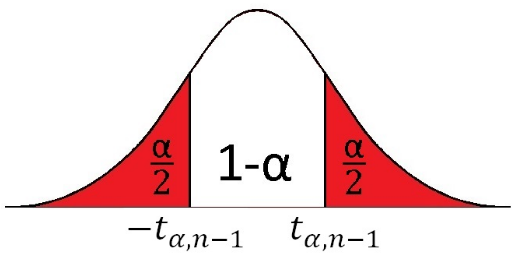

Figure 2: Student's t-distribution with critical region at significance level α.
Source: https://www.statystyka-zadania.pl/...

Student's t-test for Parameter Significance

The Student's t-test is used for testing hypotheses about a population when its standard deviation is unknown, which is typical in the analysis of regression coefficients.

t-statistic

In general, the t-statistic is calculated as:

$$t = \frac{\text{Estimator} - \text{Hypothesized Value}}{\text{Standard Error}}$$

In linear regression, where we test whether the coefficient β_j differs significantly from zero, the statistic takes the form:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

where $SE(\hat{\beta}_j)$ is the standard error of the estimated coefficient.

Testing Significance of Regression Parameters

The goal is to determine whether a given predictor X_j in the regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

has a statistically significant influence on the dependent variable Y.

Procedure for Testing the Significance of a Coefficient β_j

1. Formulate Hypotheses:

   $H_0 : \beta_j = 0$ (no linear relationship)

   $H_1 : \beta_j \neq 0$ (a linear relationship exists)

2. Compute the Test Statistic: $t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$.

3. Determine the Critical Region: Based on the t-distribution with $n - p - 1$ degrees of freedom, we calculate the quantile $t_\nu(\frac{\alpha}{2})$.

4. Make the Decision: Compare whether $|t| > t_\nu(\frac{\alpha}{2})$; if so, the result lies in the critical region and we reject H_0.

Example

Assume we estimate the model examining the effect of age on income. We estimate:

$$\text{Income} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{Age} + e$$

with $\hat{\beta}_1 = 0.8$ (meaning that an increase in age by one year is associated with an average income increase of 0.8 units), $SE(\hat{\beta}_1) = 0.3$ (the standard deviation of the estimator), $n = 30$ (number of samples). Test whether β_1 is significant using the t-distribution quantiles available here.

Task 4 [2 pts]

A medical company develops a diagnostic model for detecting a rare disease X. Disease X occurs in about 2% of the population. In a validation study of 1000 patients, 20 actually had the disease and 980 did not. The model predicted 50 positive cases, of which 15 were true and 35 were false.

	Disease X = Yes	Disease X = No
Predicted = Yes	TP	FP
Predicted = No	FN	TN

Table 1: Confusion Matrix of the diagnostic model

0.25 pts Complete the confusion matrix.

5×0.2 pts Provide the definitions/formulas of the following metrics. Calculate their values. Interpret the obtained results in the context of diagnosing a rare disease.

- Accuracy
- Precision
- Recall (Sensitivity)
- Specificity
- F1-score

0.5 pts Which metrics are most informative and why? Give other real-world examples of imbalanced datasets.

0.25 pts What accuracy can be achieved by a trivial classifier (and which classifier is it)?

Task 5 [1 pt]

Given $f(x) = x^2 + 2x$, perform two steps of gradient descent starting from $x_0 = 2$ with learning rate $\eta = 0.3$.

Discussion Topics for the Lab

- What other metrics than MSE can be used for regression?
- What is the difference between a parameter and a hyperparameter? How to choose η?
- How to split data properly in imbalanced-class problems?
- What difficulties may occur using GD in ML, and what are modern optimization methods?