Mondragon Unibertsitatea
Escuela Politécnica Superior

Uncertainty estimation

Outline

- Introduction
  - Uncertainty sources
  - Difficulties and feasibility assumptions
- Confidence intervals for model parameters uncertainty
- Confidence intervals for standard error uncertainty
- Prediction intervals for single predictions uncertainty

Introduction

Mondragon Unibertsitatea
Escuela Politécnica Superior

- A **single prediction** is a piece of information, but a limited one.
- The algorithm takes a "decision", but how certain is the algorithm about it?
- In order to answer that, we need to discuss the **sources of uncertainty**
- **<u>Statistical uncertainty</u>**: Inherent to the process and its environment. Irreducible in practice.
- **<u>Epistemic uncertainty</u>**: Coming from the fact that a model explains data, which itself explains the process. So two related deviations occur.
- Most of the applications come from Statistics and are exclusive for regression problems, because the error can be directly quantified.
- Nevertheless, in classification problems we can also estimate uncertainties

Introduction

- Uncertainty yes, but where? And how?

- **Model parameters**, e.g. *a* and *b* in simple linear regression $\hat{Y} = f(X) = a + bX$. Confidence intervals on the parameters.

- **Model general behavior**, i.e. the global error. Confidence intervals on the model error.

- **Model single prediction**. Prediction intervals on single predictions.

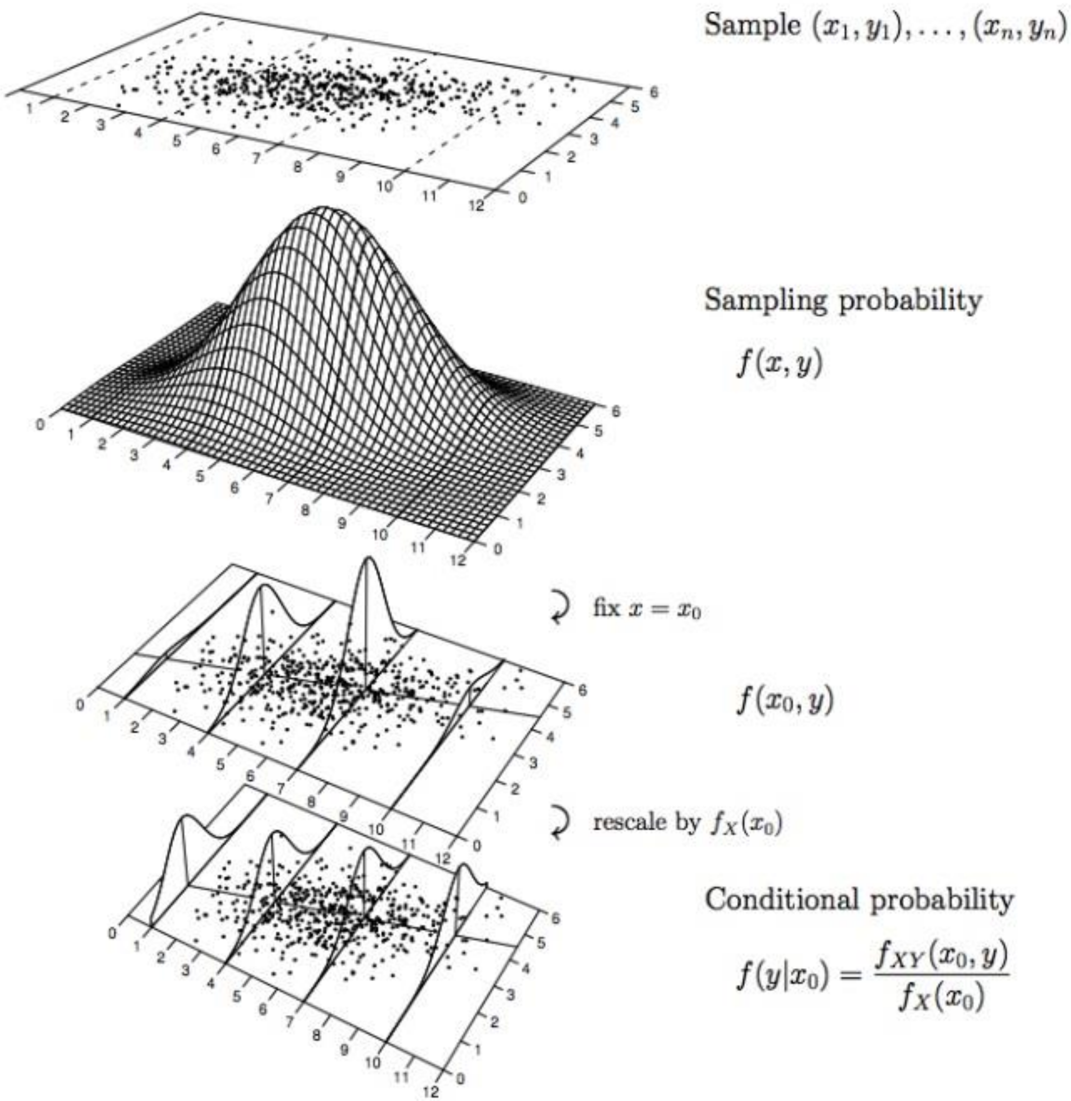
- In general, the calculations of the intervals is not simple in general. We will focus on statistical regression models, concretely on simple linear regression.

Confidence intervals for model parameters

- The most common way is by means of **bootstrapping**. The steps would be
- Create k datasets by sampling with replacement
- Train the model for all of them
- Store all k parameter collections obtained
- Use percentiles to cut the tails (usually 2.5% on each side)
- The remaining 95% of values determine the intervals (one per parameter)

Confidence intervals for standard error

Mondragon Unibertsitatea
Escuela Politécnica Superior



Prediction variance

$$\mathrm{Var}\left(\hat{\alpha} + \hat{\beta} x_d\right) = \mathrm{Var}(\hat{\alpha}) + \left(\mathrm{Var}\,\hat{\beta}\right) x_d^2 + 2 x_d \,\mathrm{Cov}\left(\hat{\alpha}, \hat{\beta}\right)$$

$$\mathrm{Var}\left(\hat{\alpha} + \hat{\beta} x_d\right) = \sigma^2 \left(\frac{1}{m} + \frac{(x_d - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)$$

Prediction standard deviation

$$s_{\mu_{y|x}} = \sqrt{Var(\hat{\alpha} + \hat{\beta} x_d)}$$

Confidence interval for standard error

$$\boxed{\mu_{y|x} \pm t_{1-\alpha/2, n-2} s_{\mu_{y|x}}}$$

Prediction intervals

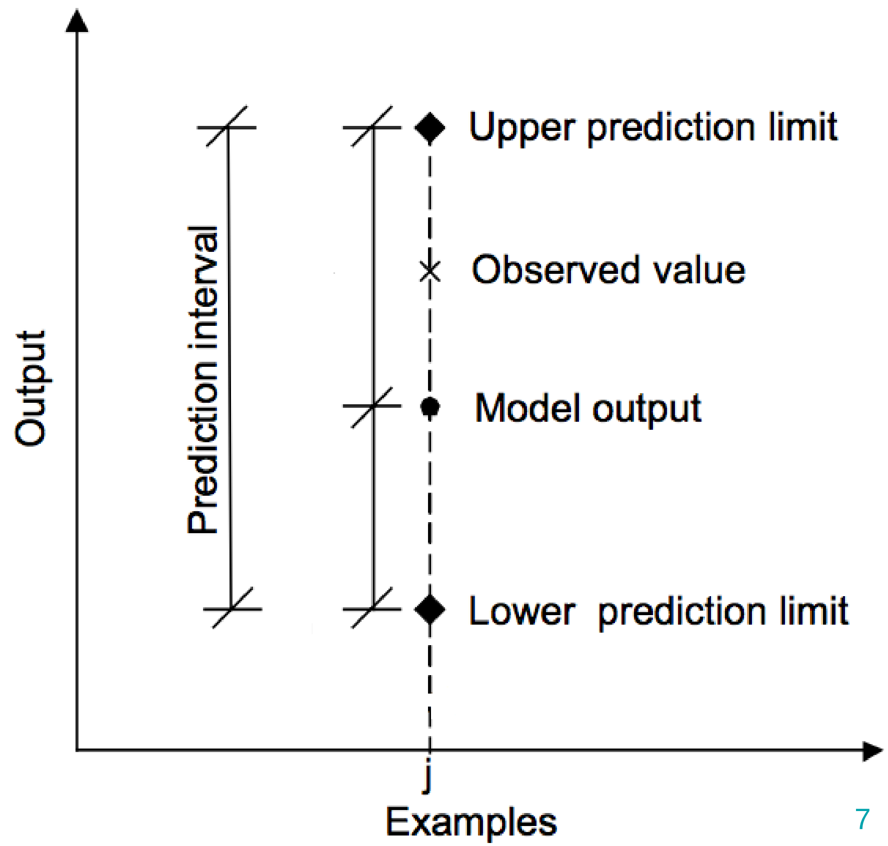
- They quantify the **uncertainty in every single prediction**

- The main difference with the confidence intervals for standard error is that they consider also the dispersion on the target, i.e. they consider that the standard deviation of prediction is

$$s_{y|x} = \sqrt{Var(y_d) + Var(\hat{\alpha} + \hat{\beta}x_d)}$$

- Because of that difference, prediction intervals are wider than confidence intervals wrt similar zones of the prediction space.

- When we plot them, they look like vertical error bars, that form an error band when plotted together.



Mondragon Unibertsitatea
Escuela Politécnica Superior

Eskerrik asko

Muchas gracias

Thank you

Carlos Cernuda
ccernuda@mondragon.edu

MGEP
Goiru, 2
20500 Arrasate – Mondragon
Tlf. 662420414