Introduction to Gaussian Processes

Part 3: Active Learning
Continuous case

Ruben Martinez-Cantin

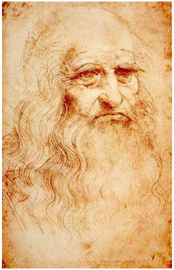
Defense University Center
Zaragoza, Spain
rmcantin@unizar.es

Outline

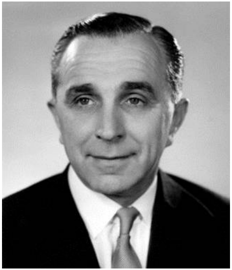
What you will see here:

- Gaussian process hyperparameters
- Regression
- Binary classification
- Active learning and experimental design
- Submodularity
- Bayesian optimization
- Stochastic bandits

A dream about flying



Leonardo



Henry Kremer

Paul MacCready

1959 Kremer had Leonardo's dream

- He offered $100 000 for the 1st human powered flight
- During 18 years, many of the best teams tried to win it.

1976 MacCready finds out about Kremers prize.

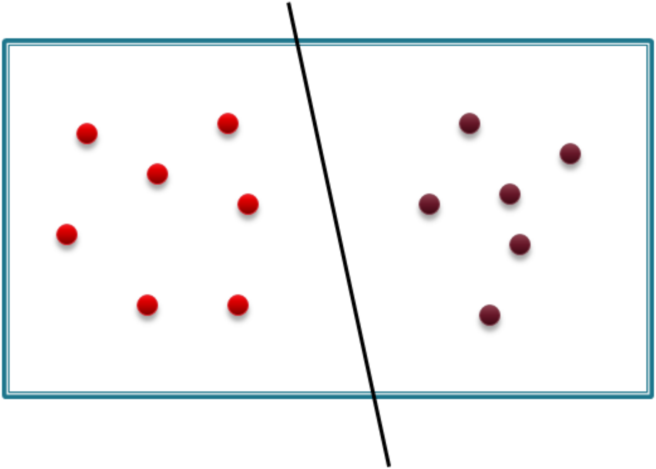
- Motivation: He had a debt of exactly $100 000
- His team had no experience on building planes

Six months later...

How did they do it?

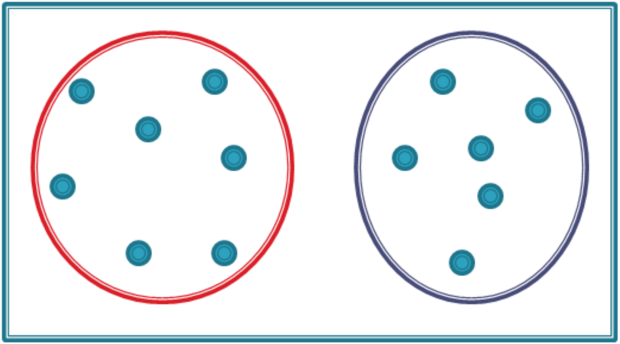
- They build a lightweight model that was easy to repear.
  - Other teams were using 1 year to build and test a model.
  - MacCready's team was able to do that in days or even hours.
  - **It is all about data gathering.**
- Their inexperience was a plus
  - Other teams were always refining their best design, doing small changes.
  - MacCready decided to forget everything he knew about aeronautics.
  - **Explore your parameter space. Do not get stuck in local minima.**

Supervised learning



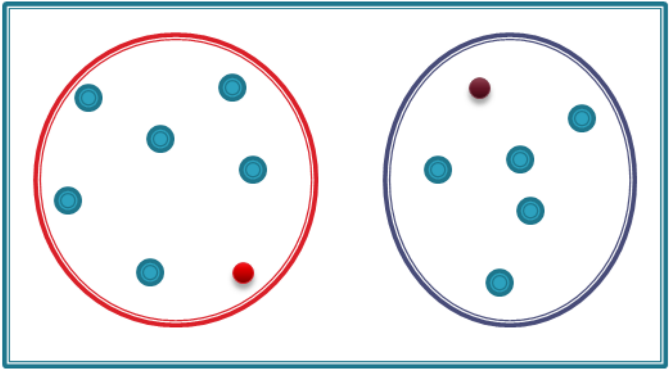
We have inputs and labels

Unsupervised learning



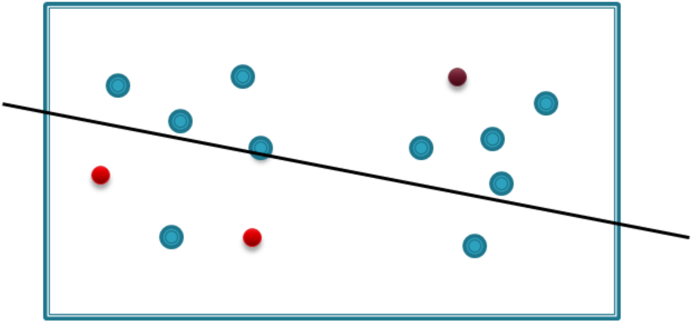
We have inputs and structure.

Semisupervised learning



We have inputs, some labels and structure.

Active learning



We have inputs, some labels, structure and we can **ask**.

Optimality

- When we need to take a decision, we need a cost, loss or regret function $\delta(f, d)$ that modulates our decisions d.
- Average case analysis:

$$d_{ac} = \arg\min_{d} \int_{F} \delta(f, d) dP(f)$$

- We have already seen this.
  - Label assignment in classification.
- Gaussian processes are an easy way to solve the integral.

Some loss examples

- Global (non-convex) optimization with budget N.

$$\delta_{GO}(f,d) = f\left(x_N(d)\right) - f(x^*)$$

- Stochastic bandits with budget N.

$$\delta_{SB}(f,d) = \frac{\sum_{n=1}^{N} f\left(x_n(d)\right) - f(x^*)}{N}$$

- A-optimality (error minimization)

$$\delta_{AO}(f,d) = (f-\hat{f})^T \Sigma (f-\hat{f})$$

- D-optimality (entropy minimization)

$$\delta_{AO}(f,d) = \mathcal{H}(f|x_n)$$

Transforming losses to criteria

Let be $p(y_t|\mathbf{x}_{1:t-1}, y_{1:t-1}, \mathbf{x}_t) = \mathcal{N}(\mu_t, \sigma_t^2)$ and $y_{min} = \inf(y_{1:t-1})$

Predicted mean μ_t

Predicted variance σ_t

Predicted mean and variance ratio $|\mu_t|/\sigma_t$.

Margin distance $|\mu_t|$

Upper or lower bounds $\mu_t \pm \beta_t \sigma_t$

Probability of improvement

$$p(y_t \leq y_{min}) = \Phi\left(\frac{y_{min} - \mu_t}{\sigma_t}\right)$$

Expected improvement

$$EI(x) = \mathbb{E}[I(x)] = \int \max(y_{min} - f(x), 0)\, dp(f)$$

$$EI(x_t) = (y_{min} - \mu_t)\Phi\left(\frac{y_{min} - \mu_t}{\sigma_t}\right) + \sigma_t \phi\left(\frac{y_{min} - \mu_t}{\sigma_t}\right)$$

DIRECT: DIvide RECTangles ⇒ Lipschitz optimization