Machine Learning
Exercise 5

Marc Toussaint
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

May 4, 2016

1 Special cases of CRFs

Slide 03:27 summarizes the core equations for CRFs.

a) Confirm the given equations for $\nabla_\beta Z(x,\beta)$ and $\nabla^2{}_\beta Z(x,\beta)$ (i.e., derive them from the definition of $Z(x,\beta)$).

b) Binary logistic regression is clearly a special case of CRFs. Sanity check that the gradient and Hessian given on slide 03:10 can alternatively be derived from the general expressions for $\nabla_\beta L(\beta)$ and $\nabla^2{}_\beta L(\beta)$ on slide 03:27. (The same is true for the multi-class case .)

c) Proof that ordinary ridge regression is a special case of CRFs if you choose the discriminative function $f(x,y) = -y^2 + 2y\phi(x)^\top\beta$.

2 Structured Output: mixed classification/regression

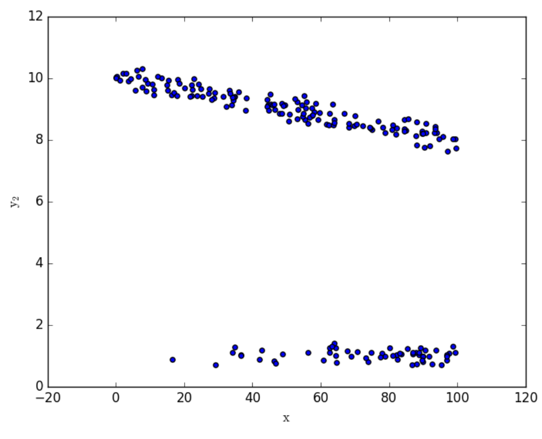
Note: The following two exercises are meant to train you in modelling structured output. The generic application case would be time series tagging or image segmentation. However, for both of these applications defining features and computing the $\text{argmax}_y f(x,y)$ is non-trivial. So, for the sake of simplicity of the exercises I describe some made up 'toy' problems.

Consider a data set of tuples (x, y_1, y_2) where

- $x \in \mathbb{R}$ is the age of a machine in some factory
- $y_1 \in \{0,1\}$ is a binary indicator whether some piece of the machine got loose or broke
- $y_2 \in \mathbb{R}$ is the performance of the machine (perhaps the rate of correct production/assembly or so).

The figure displays the data (x, y_2), which looks like two curves to be fit. These two curves correspond to $y_1 \in \{0,1\}$.

a) Properly define a *representation* and *objective* for modelling the prediction $x \mapsto (y_1, y_2)$.



b) Implement the method. What is your prediction for $x = 80$, that is, what is the probability $P(y_1 \mid x = 80)$ of a broken piece and, conditionally to that, your prediction of performance y_2 conditional to $(y_1, x = 80)$, i.e., for both cases $y_1 \in \{0, 1\}$?

3 Structured Output: multi-output regression

Consider data of tuples (x, y_1, y_2) where

- x is the age of a spotify user
- $y_1 \in \mathbb{R}$ quantifies how much the user likes HipHop
- $y_2 \in \mathbb{R}$ quantifies how much the user likes Classic

Naively one could train separate regressions $x \mapsto y_1$ and $x \mapsto y_2$. However, it seems reasonable that somebody that likes HipHop might like Classic a bit less than average (anti-correlated).

a) Properly define a *representation* and *objective* for modelling the prediction $x \mapsto (y_1, y_2)$.