Machine Learning 1

Original Machine Learning Sample Exam

Original Sample Exam from 06.02.2017

Name:

Foreword

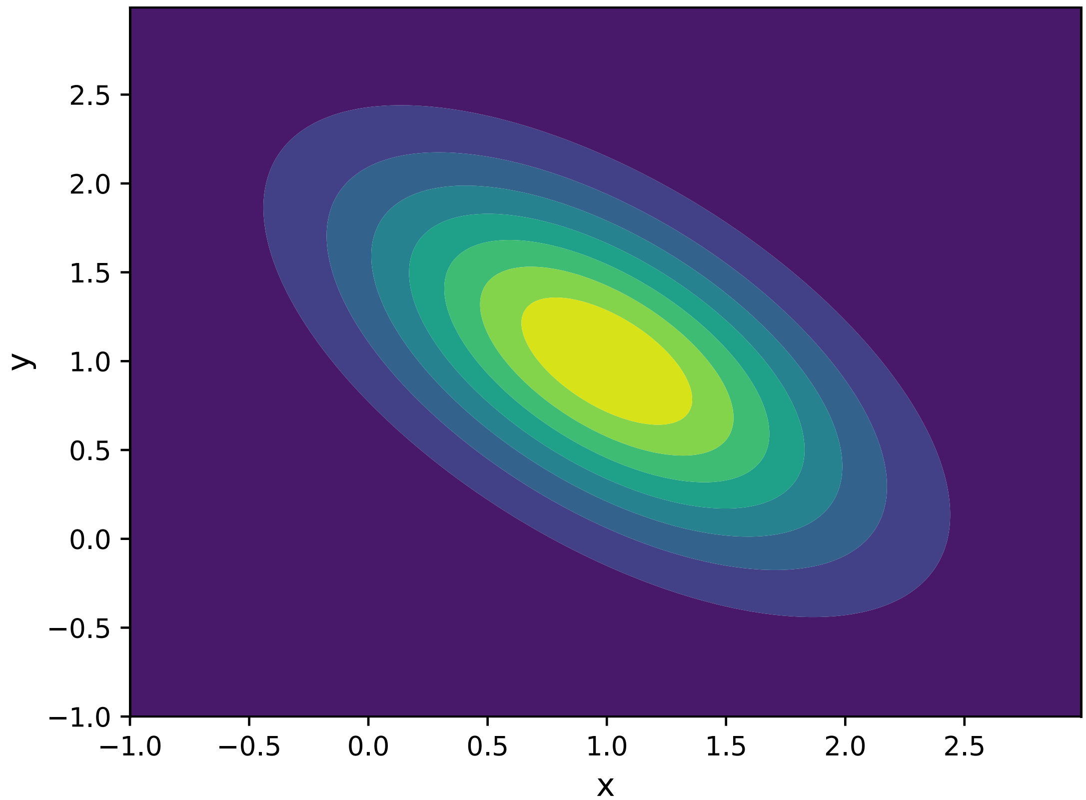
This is an official sample exam from 06.02.2017. As the original latex file has been lost, all questions and plots have been faithfully recreated. This was only possible thanks to the brave souls who secretly photographed their submissions in the post-exam review. Thank you a lot!

The questions may seem hard or even confusing at the first glance, but with practice you'll ace through them in no time!

Good luck and most importantly, have fun!
-Tristan

Q.1 (10 points)

The figure below illustrates the probability density function of a random variable (X, Y) where a bright color represents a high density and a dark color a low density. Which of the following statements is correct?



□ X and Y are uncorrelated.

□ X and Y are positively correlated.

□ X and Y are negatively correlated.

□ No statement about correlation of X and Y possible.

Q.2

Which of the following statements about likelihood functions are true and which are false? For each correct answer two points are given, for each wrong answer two points are subtracted. The total number of points for this exercise cannot fall below zero.

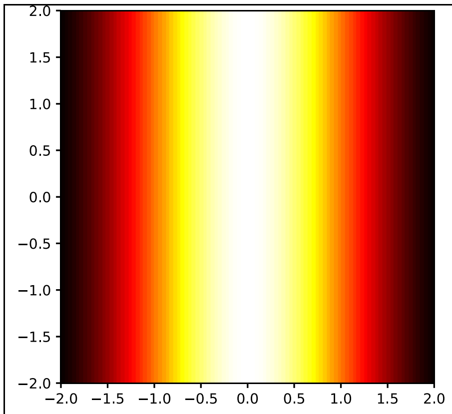
True	False	
☐	☐	Maximizing the likelihood function leads to the same estimator for the parameters as maximizing the log-likelihood function.
☐	☐	In general, the likelihood function is a probability distribution.
☐	☐	The likelihood function of the mean μ for independent and identically distributed (i.i.d.) random variables $\sim N(\mu, \sigma)$ is proportional to a normal distribution.
☐	☐	The maximum likelihood estimate for the variance σ for i.i.d. random variables $\sim N(\mu, \sigma)$ is unbiased.

Q.3

You fit a polynomial of degree one, i.e. $y(x, \mathbf{w}) = w_0 + w_1 x$ to a dataset $D = \{(x_n, t_n): i = 1, \dots, N\}$ by Bayesian linear regression. One tuple in this dataset is given by $(x_1, t_1) = (1{,}0)$.

a) How does the likelihood function $L(\mathbf{w}|x_1, t_1)$ look like? Choose from the figures below, which illustrate possible likelihood functions. A bright color indicates a great value while a dark color indicates a small value.

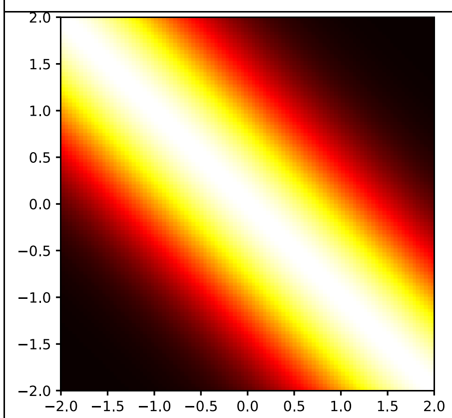
b) For each of the other plots, decide whether or not a data point (x, t) exists so that the likelihood function has the correspondent form. Specify explicitly such a point.



☐ Corresponds to $(x_1, t_1) = (1{,}0)$

☐ Corresponds to other data point: _______

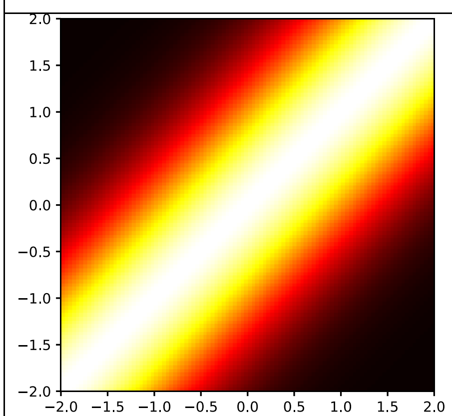
☐ No such point exists



☐ Corresponds to $(x_1, t_1) = (1{,}0)$

☐ Corresponds to other data point: _______

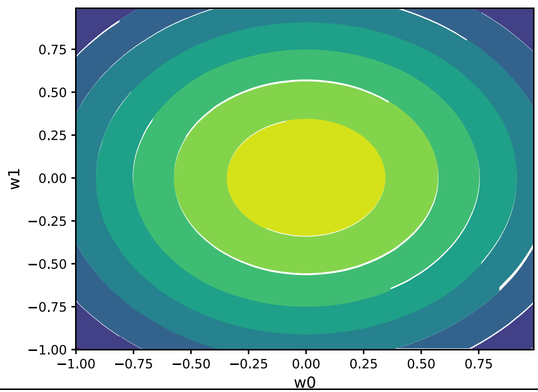
☐ No such point exists



☐ Corresponds to $(x_1, t_1) = (1{,}0)$

☐ Corresponds to other data point: _______

☐ No such point exists



☐ Corresponds to $(x_1, t_1) = (1,0)$

☐ Corresponds to other data point: _______

☐ No such point exists

Q.4

You perform Bayesian linear regression on a dataset where the target $t_n = f(x_n) + \epsilon$ is given by the sum of an unknown deterministic function f evaluated at input x_n and zero mean Gaussian noise ϵ with fixed precision parameter β for $n = 1, \dots, N$. Assume you have M basis functions $\phi_0(x), \dots, \phi(x)_{M-1}$.

a) Results increasing the number M of basis functions always in better generalization, even if the number of training samples N is limited? Answer with a short justification.

b) How does the prior change as the number of training samples N grows? Answer with a short justification.

c) How does the mean of the predictive posterior distribution, evaluated at x, depend on the basis functions?

d) Is the variance of the predictive posterior distribution independent from the input x?

e) Give the definition of the design matrix $\mathbf{\Phi}$.

Q.5

$$p(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, for\ x \geq 0 \\ 0, otherwise \end{cases}$$

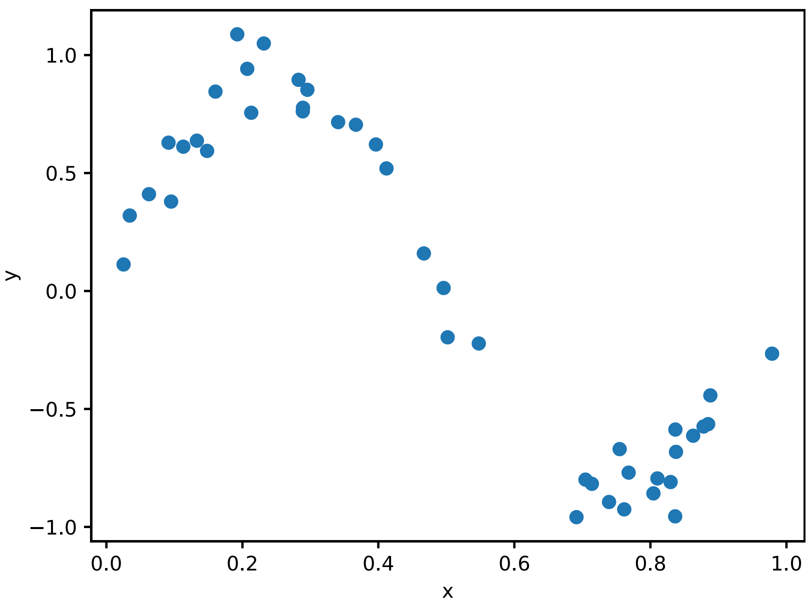
$$p(\lambda) = \begin{cases} e^{-\lambda}, for\ \lambda \geq 0 \\ 0, otherwise \end{cases}$$

a) Compute the maximum likelihood solution for the parameter $\lambda > 0$ given you observed i.i.d. samples $x_1, \dots., x_N \geq 0$.

b) Write down the explicit expression for the posterior $p(\lambda|x_1, \dots, x_N)$. You may use $p(x_1, \dots, x_N) = c$ and you do not need to simplify the expression.

Q.6

a) Which choice of basis functions (besides $\phi_0(x) = 1$) for Bayesian linear regression on the dataset below is most appropriate?



☐ $\phi_1(x) = \exp\left(-\frac{1}{2\sigma^2}(x - 0.25)^2\right)$, $(x) = \exp\left(-\frac{1}{2\sigma^2}(x - 0.8)^2\right)$ with $\sigma = 2$

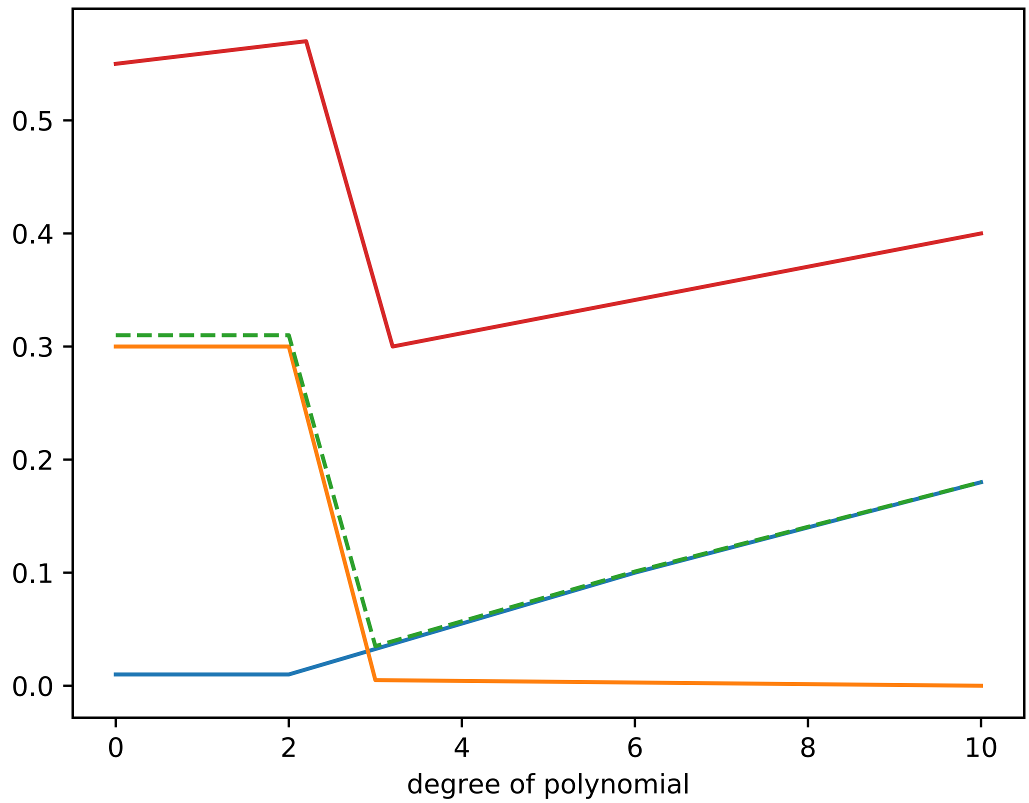
☐ $\phi_k(x) = x^k \ for \ k = 1,2,3$

☐ $\phi_k(x) = \cos(kx) \ \ for \ k = 1, \dots, 4$

b) According to the bias variance decomposition:

$$expected\ loss = (bias)^2 + variance + noise$$

The figure below illustrates squared bias, variance as well as the sum of both and the mean squared error on a testing set for models of different complexities. Here, the model is a polynomial and its complexity the degree of the polynomial which ranges from 1 to 10. Label each line in the figure and interpret them with respect to underfitting and overfitting. Which model is the best choice?



Q.7

a) Consider the dataset $D = \{(-1,2), (0,0), (1,1), (2,-1)\}$ which contains the tuple (x_n, t_n) of input x_n and output t_n. Calculate for a linear model for regression $y(x, \mathbf{w}) = (w_1 \quad w_0) = \binom{x}{1} = w_1 x + w_0$ on this dataset, the maximum likelihood solution $\mathbf{w}_{\mathrm{ML}}$.

Hint:

$$\mathbf{w}_{\mathrm{ML}} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \mathbf{t}$$

where $\boldsymbol{\Phi}$ is the design matrix.

b) Draw a sketch of the dataset as well as of the fitted curve.

c) Predict the output for $x = 9/8$.

Q.8

Is it possible to train a (successful) model for logistic regression on this dataset of tuples (x_1, x_2) split into two classes (circles and crosses) without performing a transformation $\phi(x_1, x_2)$? If not, is it possible to do so on a transformed dataset? If this is the case, write down an appropriate choice for $\phi(x_1, x_2)$. Support your answer by sketches.