Visual Computing Institute
Computer Vision
Prof. Dr. Bastian Leibe

Written Examination

2016-07-21

Name:	Mustermann	First Name:	Max
Program of Study:	Informatik Master		
Matr.-No.:	123456	Exam #:	123

Information:

- Write your name and matriculation number on **every sheet of paper**.
- Make sure that all **11 pages** are present.
- Answer each question on the provided sheet. If more space is needed, **use a new sheet of paper for each question**.
- If you have to draw to answer a question, multiple templates are provided. **Cross out wrong answers!**
- At the end of the examination this cover sheet together with the question sheets and all additionally used papers have to be returned.
- Duration of the exam: **60 minutes**.
- **No additional aids** (notes, calculator, ...) are allowed.
- Write **legibly**. Not readable text will not be graded.
- Use a pen with **blue or black ink** for writing down your solutions. Text written with pencils or red/green pens will not be graded.

With my signature I confirm that I have **read and understood** the information above.

Signature

Question:	1	2	3	4	5	Total	Grade
Points:	13	10	10	14	13	60	-
Score:							
Moderation:							

ML-Exam Number

123

Question 1: ML Basics & Probabilities: .... ($\Sigma = 13$)

(a) For data D and hypothesis H, say whether or not the following equations must always be true. Answer in yes or no. **(2 pts)**

i. $\sum_h p(H = h \mid D = d) = 1$

i. ____________________

ii. $\sum_h p(D = d \mid H = h) = 1$

ii. ____________________

iii. $\sum_h p(D = d \mid H = h)p(H = h) = 1$

iii. ____________________

iv. $p(H|D) = \dfrac{p(D|H)p(D)}{p(H)}$

iv. ____________________

(b) In probability density estimation methods, we typically have a tuning parameter which acts as a smoothing factor. For example, in case of Histograms, bin size Δ is such a tuning parameter. Smoothing can also be interpreted in terms of 'bias' and 'variance'. Provide a very brief interpretation of bias and variance in terms of smoothing. How will bias and variance change if the bin size is increased? **(3 pts)**

(c) We want to represent the probability distribution for points x_n, $n = 1, \ldots, N$ by a univariate Gaussian distribution with parameters $\theta(\mu, \sigma^2)$. Express the likelihood $p(x_n \mid \theta)$ for a single data point using the equation for the Gaussian distribution. **(1 pt)**

(d) How can we estimate the parameters of θ using Maximum Likelihood for any probability distribution function? Give the main steps of the derivation. **(4 pts)**

(e) What implicit assumption did we make in this derivation? **(1 pt)**

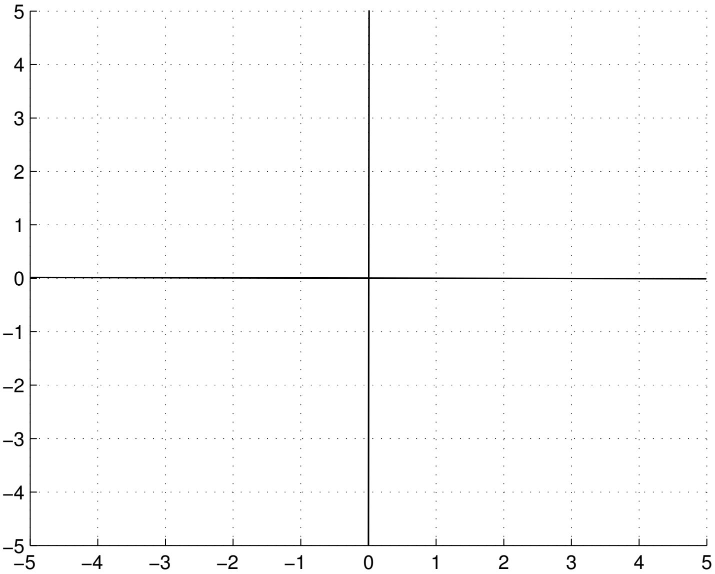
(f) What problems/limitations does Maximum Likelihood have? **(2 pts)**

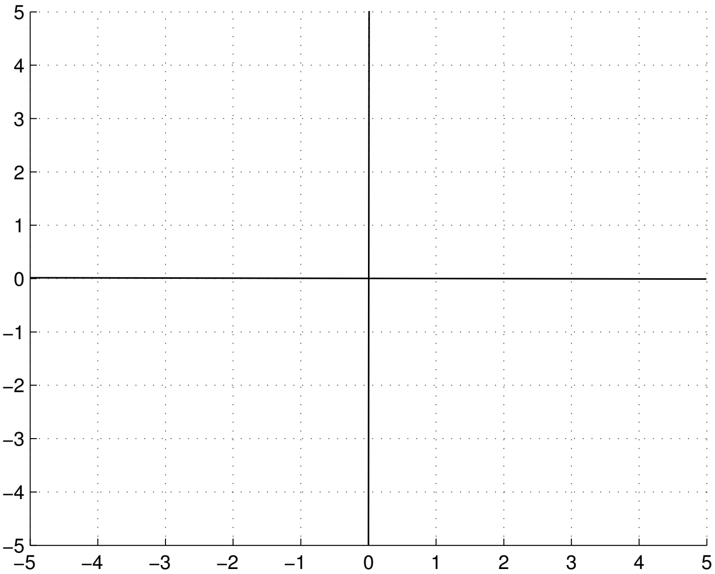
Question 2: Linear Discriminant Functions: . . . . $\left(\boldsymbol{\Sigma = 10}\right)$

(a) What is the difference between generative and discriminative methods for classification? **(2 pts)**

(b) Write an equation of the error function for a 2-dimensional, 2-class linear Least-Squares classifier and define the variables of the equation clearly. **(2 pts)**

(c) Plot the error function for the linear Least-Squares classifier. **Hint:** Use the second plot to correct your answer if needed. Strike out wrong answers. **(2 pts)**





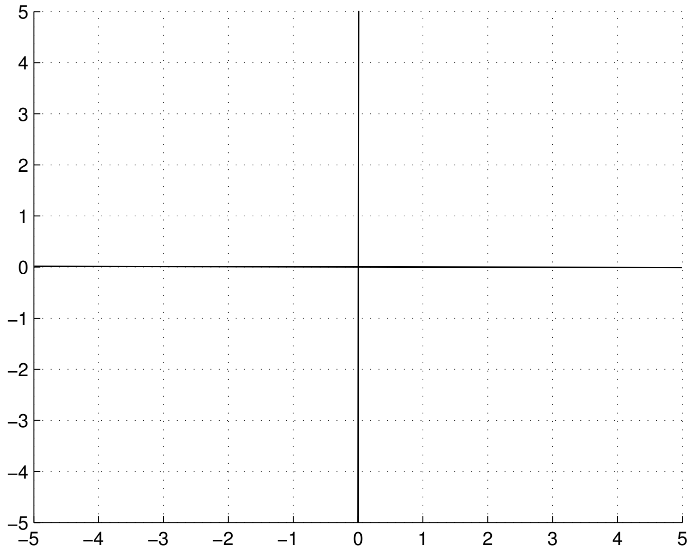
(d) Discuss the behavior of Least-Squares classification in the presence of outliers. **(2 pts)**

(e) Name two other methods that use linear discriminants and draw their corresponding error functions.

i. Name of the classification method: **(1 pt)**

i. ____________________

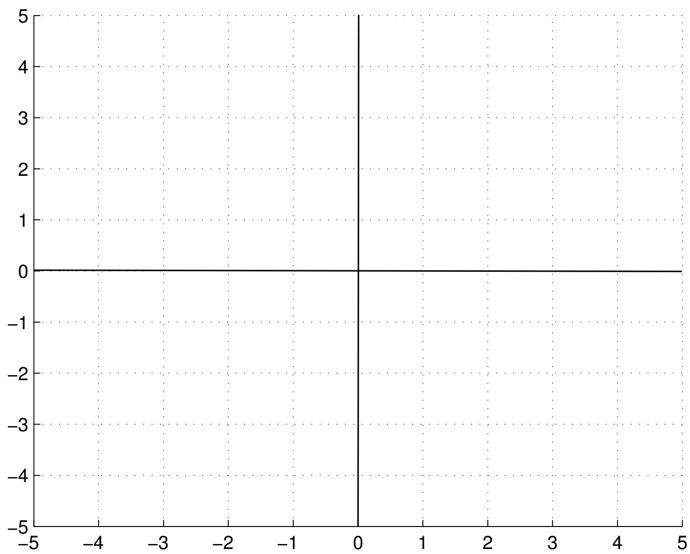
Error function:



ii. Name of the classification method: **(1 pt)**

ii. ____________________

Error function:



Question 3: VC-dimension and Support Vector Machines: .... $(\Sigma = 10)$

(a) Let H be the set of all oriented lines in the (x, y)-plane. Points on one side of the line are classified as positive and on the other side as negative. What is the VC-dimensions of H? **(1 pt)**

(b) For each of the following cases, state whether it would be best to use the primal or dual SVM formulation. Briefly explain your answer. **(4 pts)**

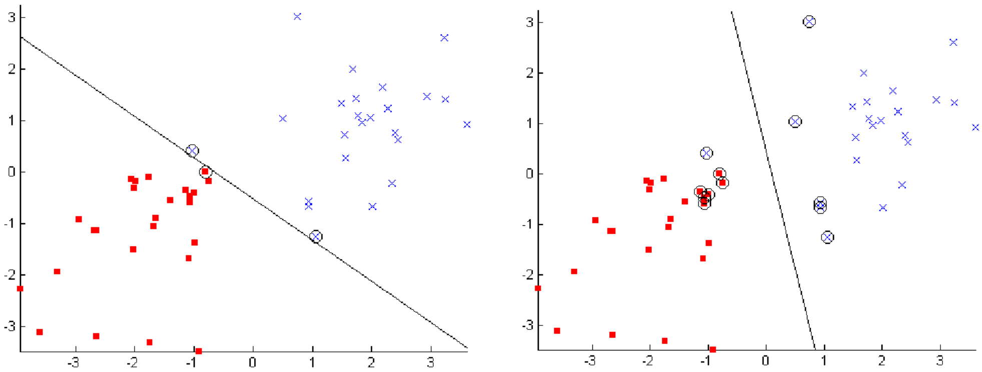
i. We apply a feature transformation that maps the input data into a feature space with infinite dimension.

ii. We apply a feature transformation that doubles the dimension of the input data. The input data has millions of training examples and is linearly separable.

(c) Is the following statement true or false? When the data is not completely linearly separable, the linear SVM without slack variables returns $\mathbf{w} = 0$. **(1 pt)**

(c) ______________________

(d) You trained a linear SVM on the toy problem below and obtained the solution shown on the left. Your friend also trained a linear SVM on the same problem, but obtained the solution shown on the right. **(2 pts)**



What happened here? How can the difference be explained?

(e) In general explain which data points will be selected as support vectors by an SVM. **(2 pts)**

Question 4: Adaboost . . . . ($\boldsymbol{\Sigma} = \mathbf{14}$)

(a) Write down the steps of Adaboost algorithm below and provide the corresponding formulas. Given a candidate pool of weak classifiers $\{h_k\}$ and training samples $\{(\mathbf{x}_n, t_n)\}$, $n = 1 \dots N$ (where $\mathbf{x}_n \in \mathbb{R}^d$ are data points in d-dimensional space and $t_n \in \{ -1, +1 \}$ are class labels), the AdaBoost algorithm for the two-class problem is:

i. Initialize the weights: **(1 pt)**

ii. For $m = 1, \dots, M$ **(6 pts)**

iii. Resulting classifier: **(1 pt)**

(b) Please state whether the statements below are true or false. Explain your answer.

i. AdaBoost will eventually reach zero training error, regardless of the type of weak classifier it uses, provided enough weak classifiers have been combined. **(2 pts)**

ii. AdaBoost can model non-linear decision boundaries. **(2 pts)**

(c) Does AdaBoost work better with strong base classifier or with weak ones? Why? **(2 pts)**

Question 5: Graphical Models .... ($\mathbf{\Sigma = 13}$)

Consider the following Bayesian Network

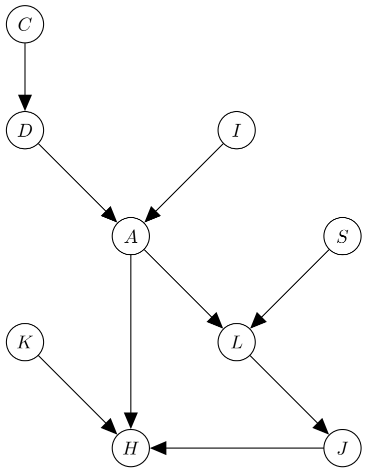


Figure 1: Bayesian Network

(a) Write down the set of variables which forms the Markov blanket of 'A' in Fig. 1? **(2 pts)**

(b) For each of the following independence assumptions, please state whether it is true or false: **(2 pts)**

i. $D \perp\!\!\!\perp S$

i. ______

ii. $D \perp\!\!\!\perp S \mid L$.

ii. ______

iii. $C \perp\!\!\!\perp J \mid H$.

iii. ______

iv. $C \perp\!\!\!\perp J \mid A$.

iv. ______

(c) Specify the factorization of the joint probability $p(A, C, D, I, S, L, J, K, H)$ of the Bayesian Network given above. **(3 pts)**

(d) Convert the given Bayesian Network in Figure 1 into an undirected graphical model and draw the resulting graph. **(3 pts)**

(e) Specify the factorization of the joint probability $p(A, C, D, I, S, L, J, K, H)$ of the resulting undirected model. **(3 pts)**