This is the formula sheet for the **Machine Learning** exam. Please print out your own copy, on a single sheet of **A4 paper**, and bring it with you. Note that most likely you will only need a few of these formulas. The rest are here for completeness, and so as not to give away the contents of the exam.

A and B are random variables. $\mathbf{x}$ and $\mathbf{y}$ are vectors. The elements of $\mathbf{x}$ are indicated by x_i.

linear algebra

dot product :
$\mathbf{x}^T\mathbf{y} = \mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i$

linear regression model :
$\mathbf{w}^T\mathbf{x} + b = y$

linear classification model :
$\mathbf{w}^T\mathbf{x} + b >^? y$

probability

joint probability :
$p(A, B)$

marginal probability :
$p(B) = \sum_a p(A = a, B)$

conditional probability :
$p(A \mid B) = \frac{p(A,B)}{p(B)}$

Bayes' law :
$p(B \mid A) = \frac{p(A|B)p(B)}{p(A)}$

univariate normal pdf :
$N(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$

multivariate normal pdf :
$N(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$

entropy

entropy (of random variable) :
$H(A) = -\sum_a p(a)\log_2 p(a)$

entropy (of probability function) :
$H(p) = -\sum_x p(x)\log_2 p(x)$

cross entropy :
$H(p, q) = -\sum_x p(x)\log_2 q(x)$

KL divergence (or relative entropy) :
$KL(p, q) = -\sum_x p(x)\log_2 \frac{q(x)}{p(x)}$
$= H(p, q) - H(p)$

Information gain :
$I_S(V) = H(S) - \sum_i {}^{|S_i|}/_{|S|} H(S_i)$

rules of derivation

constant rule $\frac{\partial c}{\partial x} = 0$

exponent rule $\frac{\partial x^n}{\partial x} = nx^{n-1}$

const. factor $\frac{\partial cf}{\partial x} = c\frac{\partial f}{\partial x}$

sum rule $\frac{\partial(f+g)}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}$

chain rule $\frac{\partial f(g)}{\partial x} = \frac{\partial f(g)}{\partial g}\frac{\partial g}{\partial x}$

gradient :
$\nabla f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]$

common derivatives

$\partial \sin(x)/\partial x = \cos(x)$
$\partial \cos(x)/\partial x = -\sin(x)$
$\partial \frac{1}{x}/\partial x = -\frac{1}{x^2}$
$\partial e^x/\partial x = e^x$
$\partial \log_b(x)/\partial x = 1/(x \ln b)$

optimization objectives

For model $f_\theta(x)$ with parameters θ, and $\text{loss}(\theta)$ as follows

least squares regression :
$\frac{1}{2}\sum_i (f_\theta(\mathbf{x}_i) - y_i)^2$

least squares classification :
$y_i = -1$ for neg. $\mathbf{x}_i$ and 1 for pos.
$\frac{1}{2}\sum_i (f_\theta(\mathbf{x}_i) - y_i)^2$

SVM classification (y_i as before) :
$\frac{1}{2}\|\mathbf{w}\| + C\sum_i \max\left(0, y_i(\mathbf{w}^T\mathbf{x} + b) - 1\right)$

SVM dual (y_i as before) :
$-\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i \alpha_i$
st. $0 \leqslant \alpha_i \leqslant C$ and $\sum_i \alpha_i y_i = 0$

logistic regression :
$y_i = 0$ for neg. $\mathbf{x}_i$ and 1 for pos.
$\text{loss}(\theta) = \sum_i H(y_i, q_i)$ with
$q_i = \sigma(\mathbf{w}^T\mathbf{x}_i + b)$

activations

sigmoid : $\sigma(x) = \frac{1}{1+\exp(-x)}$

relu : $r(x) = x$ if $x \geqslant 0$,0 otherwise

performance

accuracy : $(TP + TN)/\text{total}$

t.p.r./recall : $TP/(TP + FN)$

f.p.r. : $FP/(TN + FP)$

precision : $TP/(TP + FP)$

miscellaneous

$\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)} = \frac{\ln(x)}{\ln(2)}$
$\mathbb{E}(f(A) + g(B)) = (\mathbb{E}f(A)) + \mathbb{E}g(B)$
$\mathbb{E}(cf(A)) = c\mathbb{E}f(A)$ for const. c

The following section is your *cheat sheet*. You are allowed to write anything in the box and bring it with you into the exam, **so long as it is written by hand**. Do not write outside of the box, do not write on the back of the paper. If you violate the rules, the formula sheet will be taken away.