Australian National University

COMP4620/COMP8620

Advanced topics in AI / Foundations of AI

Theory Assignment 1

Maximum marks	75
Weight	20% of final grade
Submission deadline	Monday, 18 September 2017, 23:59
Submission mode	Via Wattle or paper or email to Sultan Majeed (max. 3MB)
Questions to	Sultan Majeed (during tutorial) or Marcus Hutter (after class)
Estimated time	20-30 hours $\approx$ 20min per mark
Late Submissions	will not be accepted, and result in 0 points, hence failure of the course
To pass the course	students must pass *each* assignment *and* the final exam

In the tutorials you have the opportunity to ask for further clarifications on the assignment questions. However, it is recommended that you give each question a serious trial beforehand. You can work and submit the assignment in groups of 2 people, provided that you work together in a collaborative way and do not split the assignment in half. By default both get the same mark but I may consider reasonable requests to apportioning the marks differently. In this case, please suggest a redistribution (e.g. move 7 marks or 5% of the marks from Alice to Bob) with a brief justification. Your names and university id must clearly be indicated in the submission. You do not need to LaTeX your solutions, provided your handwriting is readable. Start early (meaning immediately) since you likely need to prepare by spending quite some time reading and learning the material. Best split it into a sequence of mini-assignments to be solved each weak after the corresponding material has been presented in the lectures (the first 2–3 letters of the assignment label correspond to the chapter titles in the lecture slides). For logistic simplicity and your convenience we refrained from handing out 6 assignments with 6 deadlines. You should be self-disciplined enough to not do everything in the fortnight before the deadline. Please indicate the actual time spent on the assignment, and separately the time used for preparation (reading/learning).

PI-Seq (2/75) Predicting number sequences.
What is the next number in the following sequences?
(a) 1, 2, ...
(b) 1, 2, 3, 4, ...
(c) 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, ...
(d) 3, 1, 4, 1, 5, 9, 2, 6, 5, 3, ...
Can these have other continuations, other than the obvious ones?
How many numbers are enough to be sure of a continuation?

IT-Id (6/75) Identification of Strings & Natural Numbers.
(i) Every countable set is isomorphic ($\cong$) to $I\!N$ (by means of a bijection). Naively interpreting a string as a binary representation of a natural number is not unique. ($00101 \cong 5 \cong 101$). Construct some explicit (polynomial time) bijection $\langle\cdot\rangle$ between natural numbers $I\!N_0$ and finite binary strings $\{0,1\}^*$.

Unfortunately, bijections $\langle\cdot\rangle$ are not unique when concatenated,
e.g. $5 \circ 2 \cong 10 \circ 1 = 101 = 1 \circ 01 \cong 2 \circ 4$.

(ii) Develop some prefix coding for strings (called first-order), that is an injection from strings to strings ($x \to \bar{x}$) with prefix-free co-domain and $\ell(\bar{x}) = 2\ell(x) + const$.

(iii) Develop some prefix coding $x \to x`$ (called second-order)
with $\ell(x`) \sim \log_2(x) + 2\log_2\log(x)$, where $\log(\langle n\rangle) := \log(n)$.

(iv) For which x is $x`$ longer/shorter than $\bar{x}$, and how much?

(v) Write a (pseudo)algorithm that reads $\bar{x}$ from a half-infinite input stream in a self-delimiting fashion and outputs x, and similarly for $x`$.

(vi) Prove that $\{\bar{x} : x \in \{0,1\}^*\}$ and $\{x` : x \in \{0,1\}^*\}$ are prefix-free sets.

KC-KC (8/75) Properties of K.
Prove the following assertions:
(i) Lower Bound: $\sum_{x\in\{0,1\}^*} 2^{-K(x)} \leq 1$ and $K(x) \geq \ell(x)$ for 'most' x and $K(n) \to \infty$ for $n \to \infty$. Define 'most' appropriately.
(ii) Extra Information: $K(x|y) \stackrel{+}{<} K(x) \stackrel{+}{<} K(x,y)$
(iii) Subadditivity: $K(xy) \stackrel{+}{<} K(x,y) \stackrel{+}{<} K(x) + K(y|x) \stackrel{+}{<} K(x) + K(y)$
(iv) Information Non-Increase: $K(f(x)) \stackrel{+}{<} K(x) + K(f)$ for computable $f : \{0,1\}^* \to \{0,1\}^*$.
(v) For every string x, there exists a universal Turing machine U' such that $K_{U'}(x) = 1$. Argue that U' is not a natural Turing machine if x is complex.
(vi) $K(0^n) \stackrel{+}{=} K(1^n) \stackrel{+}{=} K(n \text{ digits of } \pi) \stackrel{+}{=} K(n) \leq \log_2 n + O(\log\log n)$.
(vii) The halting sequence $h_{1:\infty}$ is defined as $h_i = 1 \Leftrightarrow T_i(\varepsilon)$ halts, otherwise $h_i = 0$. Show $K(h_1...h_n) \leq 2\log_2 n + O(\log\log n)$ and $Km(h_1...h_n) \leq \log_2 n + O(\log\log n)$. [Hint: the slowest halting program among ...]

KC-Cmp (9/75) Computability.
Consider the different computability concepts: finitely computable, estimable, enumerable, (lower/upper) semi-computable, approximable=limit-computable.
(i) Show that a function is estimable if and only if it is upper *and* lower semi-computable.
(ii) Prove the implications $\Downarrow$ on slide 'Computable Functions' between the different concepts.
(iii) Show that the implications are strict, i.e. the concepts are really different.

BP-CP (4/75) Conditional probabilities.
Show that $p(\cdot|C)$ (as a function of the first argument) also satisfies the Kolmogorov axioms, if $p(\cdot)$ does.

BP-Med (4/75) Determining reliability of medical tests by Bayes' rule.
Assume the prevalence of a certain disease in the general population is 1%. Assume there exists a quite reliable test for the disease, say, the test on a diseased/healthy person is positive/negative with 99% probability. If the test (on some randomly selected person) is positive, what is the chance that (s)he has the disease? Explain the result. Hint 1: Use Bayes' rule. Hint 2: the chance is not high!

AP-RC (8/75) Relations between Complexities.
Prove in detail that
the prefix complexity $K(x) := \min_p\{\ell(p) : U(p) = x\}$,
the monotone complexity $Km(x) := \min_p\{\ell(p) : U(p) = x*\}$, and
Solomonoff's complexity $KM(x) := -\log_2 M(x) := -\log_2 \sum_{p:U(p)=x*} 2^{-\ell(p)}$
are ordered in the following way:
$0 \;\leq\; K(x|\ell(x)) - O(1) \;\leq\; KM(x) \;\leq\; Km(x) \;\leq\; K(x) \;\leq\; \ell(x) + 2\log_2\ell(x) + O(1)$

AP-CM (12/75) Convergence of M.
(i) Show that Solomonoff's bound implies that $M(0|x_{<t}) - \mu(0|x_{<t})$ tends to zero for $t \to \infty$ with μ-probability 1 for any computable μ [Hint: Use Markov inequality for $\sum_{t=1}^{\infty} a_t^2$].

(ii) Show that convergence is rapid in the sense that the expected number of times t in which $|M(0|x_{<t}) - \mu(0|x_{<t})| > \varepsilon$ is finite and bounded by c/ε^2 and the probability that the number of ε-deviations exceeds $\frac{c}{\varepsilon^2\delta}$ is smaller than δ, where $c \stackrel{+}{=} \ln 2 \cdot K(\mu)$ [Hint: Use Markov inequality for a_t^2].

(iii) Prove the instantaneous bound $M(1|0^n) \stackrel{\times}{=} 2^{-K(n)}$ and discuss its implications for sequential predictions. [Hint: Prove $1/M(0^n) = O(1)$. Use an explicit code to prove $M(0^n1) \stackrel{\times}{>} 2^{-K(n)}$. Use the K-MDL bound with $P(n) := M(0^n1)$ to prove $M(0^n1) \stackrel{\times}{<} 2^{-K(n)}$. Now combine the results.]

(iv) Show that a Universal Monotone Turing Machine with uniform random noise on the input tape outputs $x*$ with probability $M(x)$.

MDL-ML (6/75) Maximum Likelihood.
Determine an explicit expression for the maximum likely polynomial coefficients $a_{0:d}^{ML}$. [Hint: Reduce the problem to a set of linear equations, and provide a formal solution involving matrix inversion.]

MDL-Ber (8/75) MDL for Bernoulli.
Let $x = x_1x_2...x_n \in \{0,1\}^n$ be generated by a biased coin with head probability $\theta \in [0,1]$, i.e. the likelihood of x under hypothesis θ is $P(x|\theta) = \theta^k(1-\theta)^{n-k}$, where $k = x_1 + ... + x_n$.
(i) Derive an explicit expression for the ML estimate $\hat\theta := \arg\max_\theta P(x|\theta)$.
(ii) Describe a code for $\hat\theta$ and its length which upper bounds $K(\hat\theta)$.
(iii) Use $Km(x) \le -\log_2\mu(x) + K(\mu)$ with $\mu(x) := P(x|\hat\theta)$ to upper bound $Km(x)$.
(iv) Let $\tilde\theta$ be a truncated binary expansion of $\hat\theta$ to accuracy $|\tilde\theta - \hat\theta| < n^{-1/2}$.
How many bits are needed for this?
(v) Show that $0 \le \log P(x|\hat\theta) - \log P(x|\tilde\theta) \le O(1)$.
[Hint: Use the approximation $\mathrm{KL}(\hat\theta||\tilde\theta) \approx c(\hat\theta)(\hat\theta - \tilde\theta)^2$ and assume $c(\hat\theta) = O(1)$, where $\mathrm{KL}(p,q) := p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$.]
(vi) Combine the results above to the improved bound
$Km(x) \le -\log_2 P(x|\hat\theta) + \frac{1}{2}\log_2 n + O(\log\log n)$.
(Note: In general continuous classes, each parameter incurs a complexity penalty of $\frac{1}{2}\log n$ bits.)

BSP-PO (8/75) Pareto-optimality.
A distribution ξ is called *pareto-optimal w.r.t.* $\mathcal{F}$ *and class* $\mathcal{M}$ if there is no ρ with $\mathcal{F}(\nu,\rho) \le \mathcal{F}(\nu,\xi)$ for all $\nu \in \mathcal{M}$ and strict inequality for at least one ν. Show that the universal prior ξ_U is pareto-optimal for class $\mathcal{M}_U$ for total square loss $\mathcal{F}(\nu,\rho) := \sum_{t=1}^n \sum_{x_t}(\nu(x_t|x_{<t}) - \rho(x_t|x_{<t}))^2$ and for total KL loss $\mathcal{F}(\nu,\rho) := \sum_{x_{1:n}} \nu(x_{1:n}) \log\frac{\nu(x_{1:n})}{\rho(x_{1:n})}$.

The following last question is an optional (non-theory) bonus question:

USM-NCD (10/75) CompLearn Toolkit.
Reproduce the phylogenetic tree of mammals and the language tree using the CompLearn Toolkit available from http://www.complearn.org/
Describe your results and experience with this toolkit.