MIE1516 Winter 2018 Assignment 2 (10 points)

1. **Factor Encodings [2 points]:** Assume you are given the tabular representation for the conditional probability distribution: P(Z=true|X,Y) = 1/2 (X + Y) where all variables are boolean. **Provide the (a) tabular, (b) tree, and (c) canonical ADD representations for P(Z|X,Y)** (you may choose any variable ordering for trees; for the ADD, you should choose the variable ordering that gives the smallest ADD). **Explain in words why variable elimination with ADDs may be more efficient than tables for inference with P(Z|X,Y).**

2. **Metropolis-Hastings and Detailed Balance [2 points]:** Prove that *detailed balance* holds for Metropolis-Hastings with an *asymmetric* proposal distribution $q(\mathbf{x}'|\mathbf{x}^{t-1})$.

3. **Learning Markov and Conditional Random Fields [2 points]:** Find the derivative of the max conditional log likelihood of the CRF $P(\mathbf{y}|\mathbf{x}) = 1/Z(\mathbf{x}) \prod_i \psi_i(\mathbf{y}_i,\mathbf{x}_i)$ for data $D=\{(\mathbf{x}^d,y^d)\}_d$ w.r.t. parameter $\lambda_{i,j}$ where factors $\psi_i(\mathbf{y}_i,\mathbf{x}_i)=\exp\{\sum_j \lambda_{i,j} f_{i,j}(\mathbf{y}_i,\mathbf{x}_i)\}$ and $f_{i,j}(\mathbf{y}_i,\mathbf{x}_i)$ are arbitrary features over a subset of $(\mathbf{y},\mathbf{x})$. **(i)** Write the final derivative in terms of the *difference of an empirical and expected feature count* and show all derivation steps. **(ii)** Describe the computational difference between MRF and CRF learning. (Note that for an MRF, **x** is empty.)

4. **Gibbs Sampling [4 points]:** Building on your Bayesian network toolbox from Assignment 1, implement Gibbs sampling as an additional inference alternative to variable elimination.

   To test your implementation, use your graphical model from the Course Warmup and the first three queries. For **each query, show a line plot** of the expected value of the query vs. the number of Gibbs sampling time steps (blue solid line). Show the correct answer as a constant line (red dashed line). You may need to adjust burn-in or thinning for good results. **For each of the three queries, at what number of samples does the sampler appear to asymptote with the exact solution?**