Machine Learning II - Assignment 4 Anton Zeltser Harsha Nandeeshappa

Problem 1 - Binary Image Denoising - The Ising Model

Explain the role of the three factor types hxi, -8xixj, and $-\eta xiyi$. For each of the parameters, explain what happens when they are set to 0.

- -> hxi : (Unary clique potential) h controls how much importance is given to Unary prior information. that is p(x). If h is set to 0, it means that we have no unary prior info about each pixel.
- -> Bxixj: (Interaction clique potential) since xi and xj can take values -1 and +1 their product will be either -1 or +1. Since we need to minimize the energy E we need to increase the term Bxixj (because its sign is negative). Greater the ß value, more we want pixels with same sign to remain(that is we want smoother image, which help in reducing the energy). If ß is set to 0, we are not considering the smooth pixels in prior, hence the restored image can have more variation in neighbouring pixels since it is not penalised)
- -> $\eta xiyi$: This term increases if we choose the observed image(y) to be similar to restored image(x). So η controls the measure of how much the observed pixels dictate the restored image pixels. If we set η to 0, it means that we don't care how much the restored image is similar to observed image

What is the effect of setting β much larger than η ?

Value of 8 penalises the pixel xi to be different from its markov blanket(xj). Therefore setting 8 greater than η makes the image more smooth and less like its noisy observed image (y), since η penalises the image pixel(xi) if it is different from image pixel (yi from observed image).

Why does it not make sense to allow $\beta, \eta < 0$?

We need to minimize energy function E(x,y) which inturn increases p(x,y) which is proportional to (p(x|y)) Since our objective is to maximise the posterior p(x|y) if we allow $\beta, \eta < 0$, we are increasing the value of energy function and that is NOT what we want. So this does not make sense.

Problem 2 - MAP Estimation with ICM

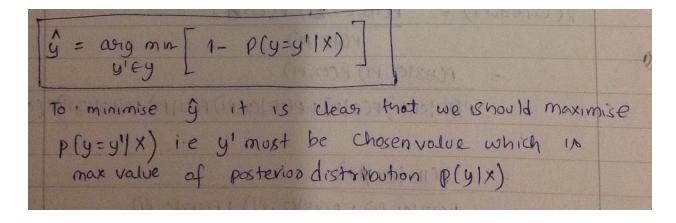
- 1. Code
- 2. Chosen parameters h = 0.8;beta = 1.2;eta = 2.1; It was not easy, had to run the program many times with different values and find it out.
- **3. h** is the least important parameter.
- 4. Tried vertical raster scan and reverse raster scan. Reverse scan took a little longer to converge. And it converged with a slightly higher error rate (3.5%). For forward vertical scan final error percentage was (3.21%)
- 5. Best result -

h = 0.800000 beta = 1.200000 eta = 2.100000 Error percentage = 4.166667 loss value = 0.041667 Error percentage = 3.267848 loss value = 0.032678 Error percentage = 3.216487 loss value = 0.032165 Error percentage = 3.210067 loss value = 0.032101 Error percentage = 3.216487 loss value = 0.032165 End

6. By Initialising the pixels at the start of the inference ICM can be improved. It could happen we happen to be at the global maxima.

Problem 3 - Sampling

1. Not sure whether it is correct (from the first assignment) I have to change the answer for p(x|y).



- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.

1.

- 1. Appearance of an intractable normalizing constant in the likelihood.
- 2. Number of non-zero entries of β (ferromagnetic coupling). Computational complexity of the log-likelihood.

2.

$$\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x}, \mathbf{y}) \Rightarrow \min_{\mathbf{x}} E(\mathbf{x}, \mathbf{y})$$

3.

4.

5.