

HOMWORK 3

MAP INFERENCE AND STRUCTURED SVM¹

10-708 PROBABILISTIC GRAPHICAL MODELS (SPRING 2021)

<http://708.mlcourse.org>

OUT: 3/10/20

DUE: 3/24/20 at 11:59 PM

TAs: Elan, Raunaq, Yiwen

START HERE: Instructions

Summary In this assignment, you will implement a Structured SVM for an image segmentation task. Section **A** will help you build a better understanding of using Integer Linear Programming for MAP inference and Loss-Augmented inference. Section **B** will guide you through the implementation of a structured SVM and you will compare its performance with a linear SVM.

- **Collaboration policy:** The purpose of student collaboration is to facilitate learning, not to circumvent it. Studying the material in groups is strongly encouraged. It is also allowed to seek help from other students in understanding the material needed to solve a particular homework problem, provided no written notes (including code) are shared, or are taken at that time, and provided learning is facilitated, not circumvented. The actual solution must be done by each student alone. The presence or absence of any form of help or collaboration, whether given or received, must be explicitly stated and disclosed in full by all involved. See the Academic Integrity Section on the course site for more information: <http://www.cs.cmu.edu/~mgormley/courses/10708/about.html#7-academic-integrity-policies>
- **Late Submission Policy:** See the late submission policy here: <http://www.cs.cmu.edu/~mgormley/courses/10708/about.html#6-general-policies>
- **Submitting your work to Gradescope:** We use Gradescope to collect PDF submissions of open-ended questions on the homework (e.g. mathematical derivations, plots, short answers). The course staff will manually grade your submission, and you'll receive feedback explaining your final marks. You will also submit your code for programming questions on the homework to Gradescope (<https://www.gradescope.com/courses/228238>). We will manually grade your code for completeness.
- For **multiple choice** or **select all that apply** questions, shade in the box or circle in the template document corresponding to the correct answer(s) for each of the questions. For \LaTeX users, replace `\choice` with `\CorrectChoice` to obtain a shaded box/circle, and don't change anything else.

¹Compiled on Tuesday 16th March, 2021 at 21:21

A Written Questions [60 pts]

Answer the following questions in the template provided. Then upload your solutions to Gradescope. You may use \LaTeX or print the template and hand-write your answers then scan it in. Failure to use the template may result in a penalty. There are 60 points and 12 questions.

A.1 Semantic Segmentation

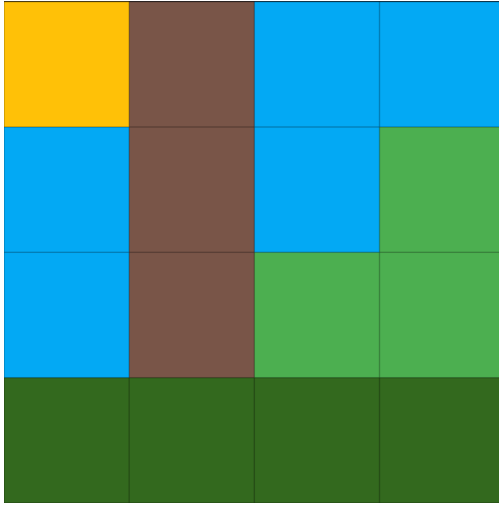


Figure A.1: Input Image

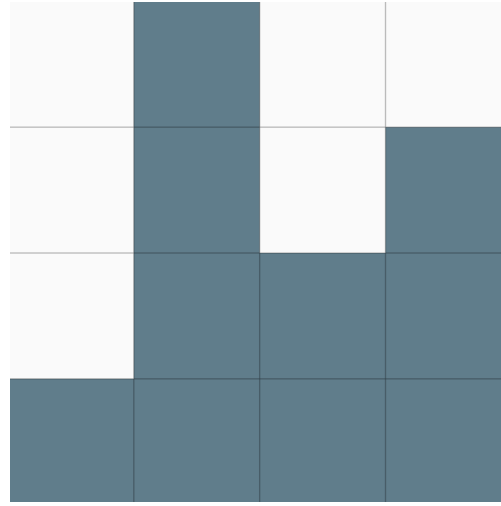


Figure A.2: Image Segmentation

Let's assume we have a 4×4 grid consisting of RGB pixels. Our task is to perform image segmentation, i.e. assign a label to each pixel from the set $\{\text{foreground}, \text{background}\}$. Intuitively, neighboring pixels should have similar values in y , i.e. pixels associated with a mountain should form one continuous blob. This knowledge can be naturally modeled via a Potts model which consists of potentials $\phi(y_i, x)$ that encode the likelihood that any given pixel is from our subject and pairwise potentials $\phi(y_i, y_j)$ which will encourage adjacent y 's to have the same value with high probability. If visualized as a graphical model, we would have a node for each pixel and an edge for each pair of pixels that are touching vertically or horizontally but not diagonally. We will refer to these sets as V and E respectively.

As seen in class, solving this task corresponds to finding y such that

$$\begin{aligned} \hat{y} &= \operatorname{argmax}_{\mathbf{y}} \log_w p(\mathbf{y}|\mathbf{x}) \\ &= \operatorname{argmax}_{\mathbf{y}} \sum_j w_j \phi(x_j, y_j) + \sum_{j,k \in E} w_{j,k} \phi(y_j, y_k) \end{aligned}$$

NOTE: This is not the same model used in the programming section. We will provide the set-up for that separately.

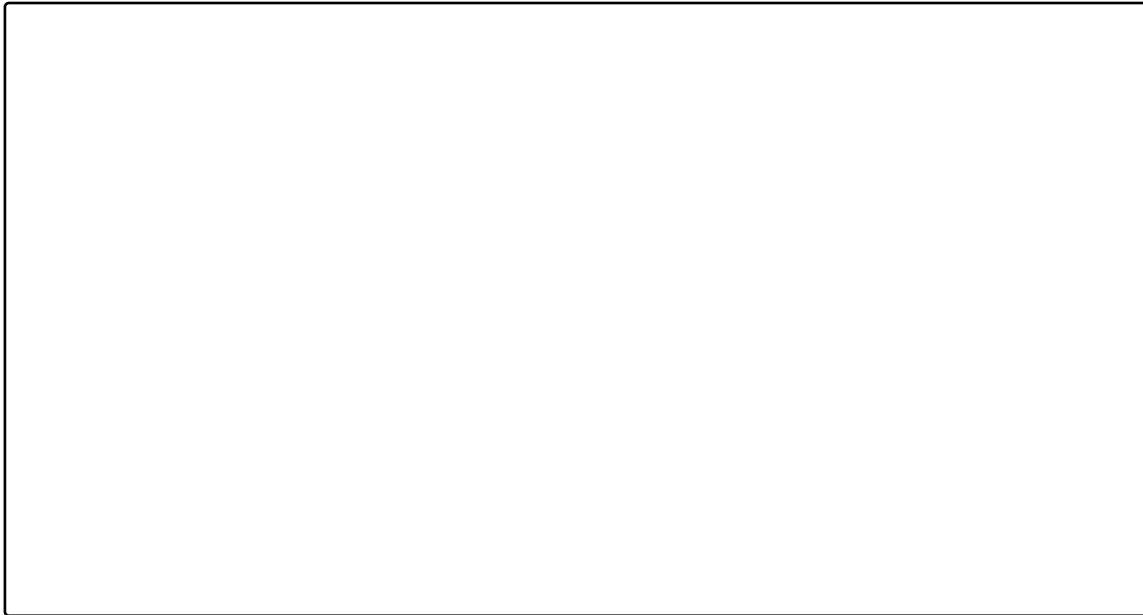
A.2 MAP Inference with Integer Linear Programming

In this section, we will formulate our problem as an Integer Linear Program (ILP) and use it to predict the segmentation of a small patch of our image. For convenience, denote the background class with 0 and the foreground class with 1. Furthermore suppose we have variables α and β that we will be optimizing over. Let $\alpha_i(y)$ represent an indicator function that is one when pixel i is labeled class y . Let $\beta_{ij}(y_n, y_m)$ represent an indicator function that is one when pixel i is set to class y_n and pixel j is set to class y_m . First we will need to translate our requirements into a constraint set.

1. (2 points) **Short answer:** Write constraints on the variables $\alpha_i(y)$ such that each pixel i is assigned one and only one class y .

2. (2 points) **Short answer:** Write constraints on the variables $\beta_{ij}(y_n, y_m)$ such that pixels i and j are assigned to the same classes y_n and y_m in the pairwise potentials as they are in the univariate functions α . Note that this will only be defined over ij pairs in our edge set E .

3. (4 points) **Short answer:** Using the score functions and goal defined above, write an ILP that represents our constraint set and objective such that if solved we would have a solution to our image segmentation task.



Suppose we had the following unary feature functions:

$$\begin{aligned}\phi_1(\text{yellow pixel, background}) &= 1 \\ \phi_2(\text{yellow pixel, foreground}) &= 2 \\ \phi_3(\text{brown pixel, foreground}) &= 3 \\ \phi_4(\text{brown pixel, background}) &= 0\end{aligned}$$

In addition suppose we had the following pairwise feature functions (note that yellow comes first, consistent with figure A.3):

$$\begin{aligned}\phi_5(\text{background, background}) &= 1 \\ \phi_6(\text{foreground, background}) &= 2 \\ \phi_7(\text{background, foreground}) &= 0 \\ \phi_8(\text{foreground, foreground}) &= 1\end{aligned}$$

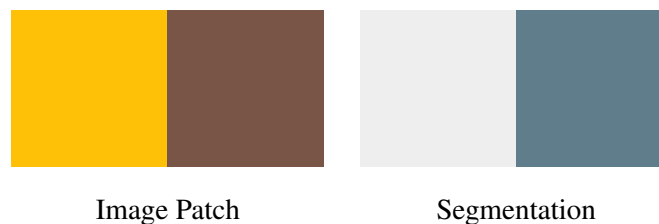


Figure A.3: Extracted Image Patch and Segmentation

4. **Short answer:** In this question we will walk through the process of solving the ILP for the 1×2 patch in the upper left hand corner of our image shown in figure A.3. Assume an initial weight vector $w = [1, 1, 1, 1, 1, 1, 1, 1]$. Without loss of generality, you can assume that a feature whose arguments are not matched evaluates to 0. For example, ϕ_5 evaluated on two foreground nodes would equal 0. Recall that the maximization problem we are trying to solve is:

$$\operatorname{argmax}_w \log p_w(y|x) = \operatorname{argmax}_w \frac{1}{Z_w} \sum_{j \in [1, \dots, 8]} w_j \phi_j(x, y)$$

- (a) (2 points) What are the scores assigned by the model for all possible y configurations shown below?

	Brown Background	Brown Foreground
Yellow Background		
Yellow Foreground		

- (b) (1 point) What is the assignment of labels to colors that solves the above ILP?

- (c) (3 points) Observe that the true label given in the figure assigns Yellow to “Foreground” and Brown to “Background”. What would be the updated weight vector w after running one iteration of structured perceptron on this label?

5. **Short answer:** Assume the same set-up as seen in question 4. Also assume an initial weight vector $w = [1, 1, 1, \dots, 1]$ and the Hamming loss as l .

- (a) (2 points) What are the loss augmented scores for all possible y configurations shown below?

	Brown Background	Brown Foreground
Yellow Background		
Yellow Foreground		

- (b) (3 points) You should observe that the loss-augmented problem has the same argmax. However, we might want instead to more heavily penalize the assignment which is “the most incorrect”, i.e. the one which assigns the opposite labels to the truth. Without modifying the features, how might we change the problem setup to achieve this on this particular prediction? Give an explicit description which will ensure the new argmax is precisely “the most incorrect” (without ties).

A.3 Loss-Augmented Inference and MBR Decoding

Suppose you are given a factor graph $\mathcal{G} = (\mathcal{C}, \psi, \mathbf{y})$ consisting of a set of factors \mathcal{C} , potential functions ψ , and variables $\mathbf{y} \in \mathcal{Y}$, where \mathcal{Y} is the output space. Each factor $c \in \mathcal{C}$ touches a subset of the variables $\mathbf{y}_c = [y_{c_1}, y_{c_2}, \dots, y_{c_m}]^T$. Further, you are given a loss function that decomposes additively according to the factors:

$$\ell(\mathbf{y}^*, \hat{\mathbf{y}}) = \sum_{c \in \mathcal{C}} \ell_c(\mathbf{y}_c^*, \hat{\mathbf{y}}_c)$$

6. (5 points) Here we consider an MRF probability distribution defined by this graph that decomposes multiplicatively according to the factors:

$$p_{\mathcal{G}}(\mathbf{y}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{y}_c)$$

Recall that the loss-augmented MAP inference problem is to find the variable assignment

$$\hat{\mathbf{y}} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \log p_{\mathcal{G}}(\mathbf{y}) + \ell(\mathbf{y}^*, \mathbf{y}).$$

Show that you can define a new factor graph \mathcal{G}' (with the same factors and variables as \mathcal{G} , but different potential functions ψ') such that the most probable assignment to $\hat{\mathbf{y}}' \triangleq \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} p_{\mathcal{G}'}(\mathbf{y})$ solves the loss-augmented MAP inference problem for \mathcal{G} with loss ℓ .

7. Now suppose you wish to build a Minimum Bayes Risk (MBR) decoder for a CRF distribution $p_{\mathcal{G}}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{y}_c, \mathbf{x})$ and loss function ℓ , where \mathcal{C} is the set of factors. For this question, you should further assume that the factor graph \mathcal{G} is acyclic and the loss function has the form:

$$\ell(\mathbf{y}^*, \hat{\mathbf{y}}) = \sum_{c \in \mathcal{C}} (1 - \mathbb{1}(\mathbf{y}_c^* = \hat{\mathbf{y}}_c))$$

In this problem you will develop a polynomial-time algorithm for computing the MBR decoder:

$$h(\mathbf{x}) = \operatorname{argmin}_{\hat{\mathbf{y}} \in \mathcal{Y}(\mathbf{x})} \mathbb{E}_{\mathbf{y} \sim p_{\mathcal{G}}(\cdot \mid \mathbf{x})} [\ell(\mathbf{y}, \hat{\mathbf{y}})]$$

- (a) (4 points) Show that the MBR decoder $h(\mathbf{x})$ can be rewritten as an argmax over a sum of factor likelihoods conditioned on the observation \mathbf{x} .

- (b) (5 points) Now describe your polynomial-time algorithm for computing $h(\mathbf{x})$ in detail. (You may assume that inference on an acyclic factor graph is polynomial-time.)



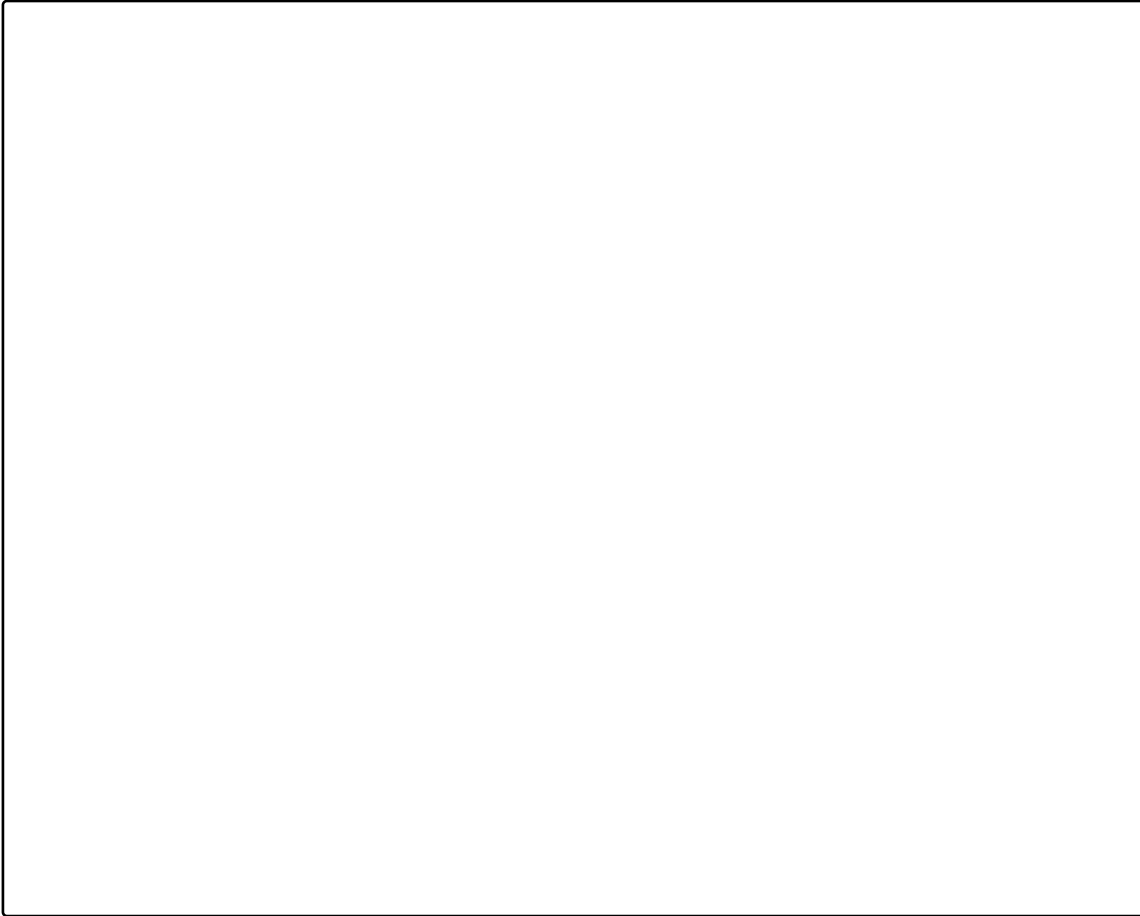
A.4 Empirical Questions

The following questions should be completed after you work through the programming portion of this assignment (Section B).

8. (12 points) Report the mean IOU and pixel accuracy scores for all models. For this question, you should train the FCN for 5 epochs and the linear SVM and structured SVM for 3 epochs each. *Round each value to two significant figures*

Model	Mean IOU	Pixel Accuracy
FCN (Train)		
FCN (Test)		
FCN+l-SVM (Train)		
FCN+l-SVM (Test)		
FCN+s-SVM (Train)		
FCN+s-SVM (Test)		

9. (10 points) Plot training and testing curves (together on one plot) for the hinge loss (FCN+l-SVM), and on a separate plot do the same for the structured hinge loss (FCN+s-SVM). 1 point per epoch. *Note: Your plot must be machine generated.*



10. (3 points) For any image of your choice from the test set, plot the labels predicted by FCN, FCN+l-SVM and FCN+s-SVM as grayscale images. Use the visualization function provided in the code to plot images. Do you see any remarkable differences between the model predictions?

A.5 Wrap-up Questions

11. (1 point) **Multiple Choice:** Did you correctly submit your code to Gradescope?

☐ Yes

☐ No

12. (1 point) **Numerical answer:** How many hours did you spend on this assignment?.

A.6 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the collaboration policy questions detailed in the Academic Integrity Policies for this course.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details including names of people who helped you and the exact nature of help you received.

2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details including names of people you helped and the exact nature of help you offered.

3. Did you find or come across code that implements any part of this assignment? If so, include full details including the source of the code and how you used it in the assignment.

B Programming [35 pts]

Your goal in this assignment is to implement a structured SVM on top of a convolutional neural network for image segmentation. Given an image as a 32x32 grid of pixels, you will assign a label from the set $\{foreground, background\}$ to each pixel.

Your solution for this section must be implemented with **PyTorch** and **Google OR-Tools**² using the data files we have provided to you. This restriction is because we will be grading your code by hand to check your understanding as well as your model's performance.

B.1 Task Background

Semantic segmentation is the task of assigning a class label to each pixel in an image. This task can be thought of as image classification, but at the pixel level. Figure B.1 shows a sample image with its corresponding segmented image.



Figure B.1: Sample image from the PASCAL VOC-2012 Image Segmentation dataset

Semantic segmentation is useful for a variety of applications such as autonomous vehicles, photo-editing tools and many tasks in robotics. In this question, we will first train a simple semantic segmentation model using a fully convolutional network (FCN). Next, we will use this FCN as a feature extractor to generate 100-dimensional features for each pixel in the image. Using these pixel features, we will build two semantic segmentation models. The first baseline model is a simple linear support vector machine (SVM) which uses the FCN features and predicts a class label for each pixel in the image independently. The second model is a structured SVM which takes in FCN features and predicts class labels for all pixels in the image while incorporating interactions between neighboring pixels. Following subsections provide more details about the dataset, metrics and model structure.

B.2 Data

For this task, we will be using the semantic segmentation dataset from the PASCAL VOC 2012 Challenge.³ This dataset consists of 2913 examples in total (2330 for training and 583 for testing). For each example in this dataset, we have a pair of images: the original image and the segmented image. To make it easier to work with the data, we have converted segmented images into numpy arrays of class labels. To ensure reasonable runtime, we have also cropped and downsampled all images to 32x32 pixels. Finally, instead of having multiple class labels (eg: person, aeroplane etc.), we will treat this as a binary classification task with the label set $\{foreground, background\}$.

²<https://developers.google.com/optimization>

³<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html>

B.3 Metrics

To evaluate model performance on this dataset, we will be using two metrics:

- **Pixel Accuracy:** Pixel-level classification accuracy for each image computed as

$$\frac{\#CorrectPixels}{\#TotalPixels} \quad (B.1)$$

- **Mean IOU (Intersection-over-Union):** Mean of per-class IOU (intersection-over-union) across all classes for each image. For one class (eg: foreground), IOU is computed as

$$\frac{\#CorrectForegroundPixels}{\#GoldForegroundPixels + \#PredictedForegroundPixels - \#CorrectForegroundPixels} \quad (B.2)$$

We will average these metrics across all images

B.4 Fully Convolutional Network

As a first step towards building a semantic segmentation model, we will train a fully convolutional network. A fully convolutional network differs from a regular convolutional neural network in the later half of its architecture. While a convolutional neural network returns a dense representation for each image (eg: a 6x6 representation for a 32x32 image), a fully convolutional network returns a representation of the same size as the original image. FCNs use regular convolution layers to first compress an image into a dense representation before *deconvolving* the dense representation to get a tensor of the same size as the original image (as shown in figure B.2).

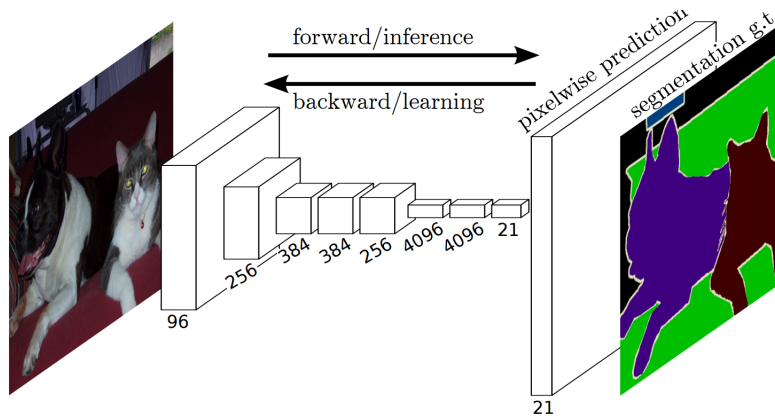


Figure B.2: Fully convolutional network for semantic segmentation

For this assignment, we have provided you with code to train and test a FCN. Our FCN implementation uses the first 5 layers from AlexNet as convolution layers and an interpolation layer to perform the deconvolution. After running the FCN on a 32x32 image, you should be able to extract a 32x32x100 tensor which contains a 100-dimensional feature vector for each pixel. We add an additional 100x2 linear layer to predict (foreground, background) scores for each pixel and train this network using cross-entropy loss. After training, we discard the linear layer and use the FCN weights to generate pixel-level features for a given image.

B.5 Baseline

The first model you will be implementing is a simple linear support vector machine. This model takes as input a $32 \times 32 \times 100$ tensor computed by the FCN and uses weights w_{fc} to calculate per-class scores for each pixel independently. This is analogous to simply adding a linear layer on top of the FCN. However, since the baseline is a linear SVM, it is trained using hinge loss. (**Hint:** To implement hinge loss in PyTorch, you might find it helpful to look up `MultiMarginLoss()`).

B.6 Main Model

The second model you will be implementing is a structured support vector machine. Unlike the linear SVM, this model does not predict labels for each pixel independently. This is useful because adjacent pixels tend to belong to the same class, indicating that there may be some benefits to treating this problem as a structured prediction problem and leveraging pixel interactions. In this assignment, we will model the structure of the output using a formulation based on associative markov networks.⁴ Associative markov networks are a class of undirected graphical models where each node is a discrete variable and clique potentials tend to favor the same label for all nodes in the clique. In our task, we treat each pixel as a node and only model pairwise interactions between pixels (i.e no cliques larger than 2 exist). For tractability, we further restrict pairwise interactions to non-diagonal neighboring pixels (up, down, left, right). This results in a lattice structure as shown in the figure:

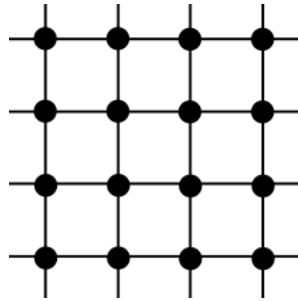


Figure B.3: Output structure used by structured SVM

Given an input image as a $32 \times 32 \times 100$ feature tensor T extracted from the FCN, a SVM over this structured output space can be formulated as follows:

- **Variables:** $x_{i,c}$ where $i \in [32 * 32]$, $c \in \{0, 1\}$, $y_{i,j,c}$, where $i \in [32 * 32]$, $j \in [32 * 32]$, $c \in \{0, 1\}$. $x_{i,c}$ and $y_{i,j,c}$ are pixel assignment and edge assignment variables respectively. Here, $[n]$ means the set of the first n natural numbers: $\{0, 1, \dots, n-1\}$. $x_{i,c} = 1$ indicates that pixel at position i belongs to class c , while $y_{i,j,c} = 1$ indicates that pixels in neighboring positions i and j both belong to class c and are linked to one another.
- **Weights:** \mathbf{w}_c^p and \mathbf{w}_c^e , where $c \in \{0, 1\}$ represents the class. For each c , \mathbf{w}_c^p , \mathbf{w}_c^e are the 100-dimensional vector pixel-weights and 200-dimensional vector edge-weights, respectively. These weights can be used to compute unary pixel log-potentials and edge log-potentials as $\phi_{i,c} = \langle \mathbf{w}_c^p, T(i) \rangle$ and $\phi_{i,j,c} = \langle \mathbf{w}_c^e, \text{concat}(T(i), T(j)) \rangle$. Recall that we used a similar process in homework 2 to compute unary/ edge potentials for nodes in the parse tree from LSTM hidden states.

⁴<https://ai.stanford.edu/~koller/Papers/Taskar+al:ICML04.pdf>

- **ILP formulation for MAP inference:** Using the pixel and edge log-potentials computed using weights and FCN features, the highest scoring output structure (i.e. the assignment of values to $x_{i,c}, y_{i,j,c}$) is chosen by maximizing the following objective:

$$\max_{\mathbf{x}, \mathbf{y}} \sum_i \sum_c \phi_{i,c} x_{i,c} + \sum_i \sum_j \sum_c \phi_{i,j,c} y_{i,j,c}$$

subject to the following constraints:

$$\begin{aligned} \sum_c x_{i,c} &= 1 \\ y_{i,j,c} &\leq x_{i,c} \\ y_{i,j,c} &\leq x_{j,c} \\ y_{i,j,c} &\geq x_{i,c} + x_{j,c} - 1 \end{aligned}$$

The first constraint ensures that each pixel is assigned to exactly one class, while the next three constraints ensure that if a link exists between two pixels, they both belong to the same class.

- **Loss-augmented Inference:** To define loss-augmented inference, we need a loss function. We choose $\ell(\hat{\mathbf{x}}, \mathbf{x}^*)$ to be the Hamming loss between the two outputs:

$$\ell(\hat{\mathbf{x}}, \mathbf{x}^*) = \frac{1}{|i||c|} \sum_i \sum_c \mathbb{I}\{\hat{x}_{i,c} \neq x_{i,c}^*\}$$

Setting up an ILP for loss-augmented inference uses the exact same constraints as the MAP Inference ILP above, but we change the objective function to be the following:

$$\max_{\mathbf{x}, \mathbf{y}} \sum_i \sum_c \phi_{i,c} x_{i,c} + \sum_i \sum_j \sum_c \phi_{i,j,c} y_{i,j,c} + \frac{1}{|i||c|} \sum_i \sum_c (x_{i,c}^* (1 - x_{i,c}) + (1 - x_{i,c}^*) x_{i,c})$$

- **Loss function:** The structured hinge loss for this SVM is as follows:

$$L = \max(0, S(\hat{\mathbf{x}}, \hat{\mathbf{y}}, T) + \ell(\hat{\mathbf{x}}, \mathbf{x}^*) - S(\mathbf{x}^*, \mathbf{y}^*, T))$$

where the scoring function $S(\mathbf{x}, \mathbf{y}, T) = \sum_i \sum_c \phi_{i,c} x_{i,c} + \sum_i \sum_j \sum_c \phi_{i,j,c} y_{i,j,c}$ is the same as the ILP objective (note that the log-potentials ϕ in the objective are computed using vectors from T), and $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is the highest scoring structure returned by loss-augmented inference, and $(\mathbf{x}^*, \mathbf{y}^*)$ is the ground truth assignment to pixel/edge variables.

Putting everything together, given an image I , you should run the following procedure:

1. Extract feature tensor $T = FCN(I)$
2. Compute pixel and edge log-potentials $\phi_{i,c}, \phi_{i,j,c}$ using T
3. Perform loss-augmented MAP inference via ILP to get highest scoring structure $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$. **Be sure to run the ILP on numpy arrays separate from the computation graph. This will ensure that PyTorch does not try to backpropagate through the ILP solver.**
4. Compute structured hinge loss with $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$

5. Backpropagate through the structured loss to update weights

At test time, you only need to return the most probable assignment as your output (i.e. step 3 below does not augment with the loss).

1. Extract feature tensor $T = FCN(I)$
2. Compute pixel and edge log-potentials $\phi_{i,c}, \phi_{i,j,c}$ using T
3. Perform MAP inference via ILP to get highest scoring structure (\hat{x}, \hat{y})

B.7 Implementation Details

We have provided you with some starter code (`starter_code.py`). You should be able to simply run this starter code to train and test a FCN on the data after adding the correct paths. You are expected to implement both SVM models by filling in the TODOs. The reference SVM implementation uses Adam optimizer with learning rate 0.0001 for both linear and structured SVM. It also uses class weights for the linear model (similar to the FCN) to handle the class imbalance issue. You should train the FCN in isolation for 5 epochs before using it as a feature extractor for any of the SVM models. Both SVM models only update their own weights and DO NOT update the FCN weights. **Since you will not be updating the FCN weights, it would be a good idea to save the image representations after training and use them as a new dataset for training the SVMs—that way, you won't have to re-evaluate the features for each iteration (make sure if you do this you still *shuffle the representations* for training the SVMs!). It would also save time to save the FCN weights and load them from a file, but **do not submit code which reads from an external file**. That is, you can do this for debugging purposes, but be sure that the code you submit trains the FCN from scratch.** Train both SVM models for 3 epochs each before answering the empirical questions. For ILP optimization, use the **CBC Mixed Integer Programming solver** from OR-Tools. If you already began writing your code using the GLOP solver, this is an easy modification: you should instantiate your solver with `CBC_MIXED_INTEGER_PROGRAMMING` instead of `GLOP_LINEAR_PROGRAMMING`, and your variables should be created as `BoolVars` instead of `NumVars` restricted to $[0, 1]$. Everything else is exactly the same!

In a previous version, we told you to use the GLOP linear solver, and noted that for binary classification the linear program will have integer solutions if the solution is unique. For our setup, this is **not true, and hence we've updated the handout to tell you to use the MIP solver. If you'd like to understand this in more detail, you can search for the post on Piazza or come to office hours!**

Here are the library versions used by one of the TAs when running this code:

```
numpy==1.19.2
torch==1.7.1
torchvision==0.8.2
ortools==8.2.8710
scikit_image==0.18.1
skimage==0.0
```

B.7.1 A Few Additional Notes

- Though usually an SVM would be trained with both the `MultiMarginLoss` and some sort of regularization (e.g., `weight_decay` in the optimizer), don't worry about the regularization for the implementation.

- Notice that the provided code for training the FCN does not equally weight the classes—this is because background pixels are much more prevalent than foreground pixels. When training the Linear SVM, you should use the same weighting scheme for the `MultiMarginLoss`. You'll notice that this technically reduces pixel accuracy, but it increases mean IOU.