Exam in SSY097

June 10th, 2019

Allowed materials : Pen/pencil, eraser.

The exam consists of six problems. Make sure that you have them all.

- Motivate all answers carefully.
- Use a new paper for each new numbered problem.
- Write on one side of the papers only.
- Write your anonymous number on each new page.
- Avoid using a red pen.
- If you want the result registered as SSY096, write this on the cover page.

The dates for the exam review will be announced on PingPong.

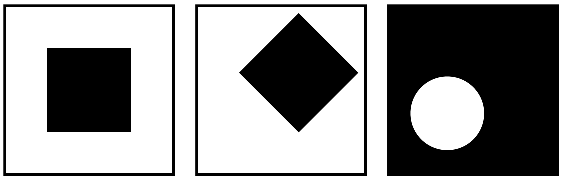
Grades

$\geq$ **8 points** Grade: 3

$\geq$ **11 points** Grade: 4

$\geq$ **14 points** Grade: 5

1 SIFT, 3 points



P Q R

(a) Consider the three patches `P`, `Q`, and `R` shown above. For each patches, draw its vector bouquet representation.

Explicitly state how many entries your bouquets have.

How many entries does a normal SIFT descriptor have? Explain how you get to your solution.

(b) In order to be robust against scale changes, SIFT builds a scale space representation and detects its keypoints as local extrema of that representation. Explain how ...

- ... the scale space representation is constructed,
- ... how the scale of each detected keypoint is computed, and
- ... how the size of the patch used to compute the SIFT descriptor is determined from the position and scale of a keypoint.

It is sufficient to describe the three steps in 2-3 sentences each, there is no need to explicitly derive / use equations.

(c) Consider the task of descriptor matching between two images $\mathcal{P}$ and $\mathcal{Q}$. The standard approach searches for each descriptor p from $\mathcal{P}$ for the nearest neighboring descriptor q_1 in $\mathcal{Q}$. However, many of the resulting matches will be wrong. In practice, it is thus common to use Lowe's ratio test[1] to filter out wrong matches.

Write down the formula for the ratio test. Make sure to explain additional variables that you introduce.

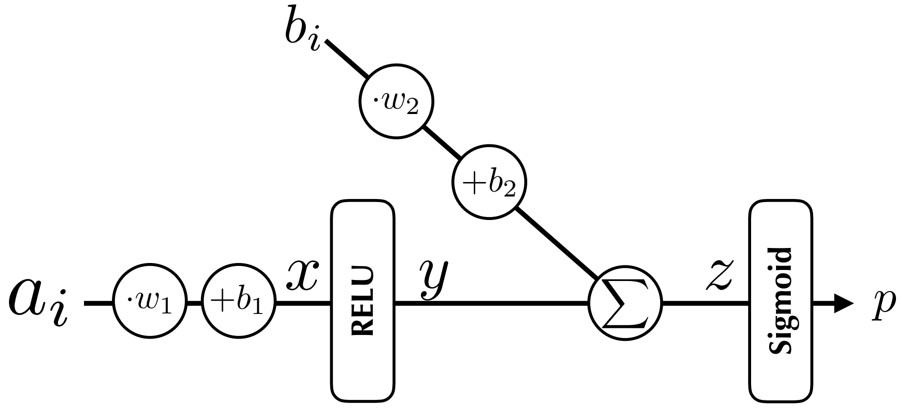
Why is the ratio test used instead of a threshold on the absolute descriptor distance $||p - q_1||_2$ to filter out wrong matches?

[1] Also known as the SIFT ratio test.

2 Statistical Learning, 3 points

Consider the neural network shown below that performs binary classification on a tuple (a_i, b_i) of two scalar input values $a_i, b_i \in \mathbb{R}$. The RELU and Sigmoid functions are given as $\text{RELU}(x) = \max(0, x)$ and $\text{Sigmoid}(z) = \frac{e^z}{1+e^z}$. The derivative of the RELU function is given as

$$\frac{\partial \text{RELU}(x)}{\partial x} = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{otherwise} \end{cases} .$$



(a) Provide the loss function L_i for a positive example (a_i, b_i). The loss should depend on the probability p estimated by the Sigmoid layer. As in the lecture and the lab, use the negative log-likelihood loss.

(b) Given a positive example $(a_i, b_i) = (10, -5)$, compute the derivatives

$$\frac{\partial L_i}{\partial w_1}, \quad \frac{\partial L_i}{\partial b_1}, \quad \frac{\partial L_i}{\partial w_2}, \quad \frac{\partial L_i}{\partial b_2}$$

through the backpropagation algorithm. To this end, first perform a forward pass to compute values for x, y, z, and p. Next, use the chain rule to derive formulas for the derivatives in the backward pass. Compute the actual values for the derivatives using your equations and the values for x, y, z, and p computed during the forward pass.
The current values for w_1, b_1, w_2, and b_2 are

$$w_1 = 10, \qquad b_1 = 50, \qquad w_2 = 5, \qquad b_2 = -125 .$$

(c) Using a learning rate of 0.01, what is the next value for w_1?

3 Robust Model Fitting & Statistical Learning, 3 points

In this question, we consider the problem of fitting a polynomial into a set of 2D data points: Given a set of N 2D points $\{(x_i, y_i) | i \in [1, \cdots, N]\}$, we are looking to estimate the parameters $\mathbf{w} = (w_0, ..., w_M)$ of a polynomial $y(x|\mathbf{w}) = \sum_{j=0}^{M} w_j \cdot x^j$ of degree M. We denote the residual between the observed y-coordinate y_i and the predicted y-coordinate $y(x_i|\mathbf{x})$ for the i-th datapoint as $r_i(\mathbf{w}) = y_i - y(x_i|\mathbf{w})$.

(a) A standard approach to estimate the parameters $\mathbf{w}$ of the polynomial is to minimize a sum of squared errors:

$$\min_{\mathbf{w}} \sum_{i=1}^{N} (r_i(\mathbf{w}))^2 \quad .$$

Show that minimizing the sum of squared errors is the same as maximizing the likelihood

$$\max_{\mathbf{w}} \prod_{i=1}^{N} p(r_i(\mathbf{w})) \quad ,$$

where $p(x)$ is a Normal / Gaussian distribution with zero mean and a standard deviation of σ (the actual value of σ is not important here).

(b) The problem of assuming that the residuals follow a Normal distribution (and thus least-squares fitting) is that a Normal distribution cannot model outliers. Consequently, outlier measurements will distort the estimated parameters.

Explain how robust cost functions can be used to reduce the impact of outliers.

Name one robust cost function discussed in the lecture (not RANSAC).

(c) In a slight modification of the problem setting from above, we are given a set of N 2D points $\{(x_i, y_i) | i \in [1, \cdots, N]\}$ that can be used for training and an additional set of K 2D points $\{(x_i, y_i) | i \in [1, \cdots, K]\}$. We still want to fit a polynomial function $y(x|\mathbf{w})$ to the datapoints, but do not know which degree M to use.

Describe an approach that determines a suitable degree M and fits the polynomial function $y(x|\mathbf{w})$ to the data while trying to avoiding overfitting.

4 RANSAC, 3 points

(a) Consider two minimal solvers for the same problem with the following properties:

- **Solver 1**: The solver requires $m = 3$ data points, takes 1 time unit to run, always returns 4 solutions[2], and evaluating one model on a single data point requires 0.01 time units.
- **Solver 2**: The solver requires $m = 4$ data points, takes 1 time unit to run, returns 1 solution if all data points are inliers and 0 solutions otherwise, and evaluating one model on a single data point requires 0.01 time units.

Derive equations for the average run-time of RANSAC with each solver for a given inlier ratio ε and N data points.

Hint: For a given inlier ratio ε and a solver that needs to sample m data points in each iteration, RANSAC on average requires $\frac{1}{\varepsilon^m}$ iterations to find an all-inlier sample.

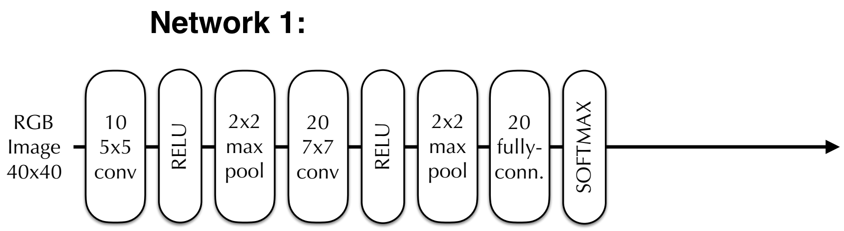
(b) Using your equations from (a), determine which solver is on average faster for $\varepsilon = 1$ and $N = 100$, $\varepsilon = 0.1$, and $N = 100$, and $\varepsilon = 0.1$ and $N = 1000$.

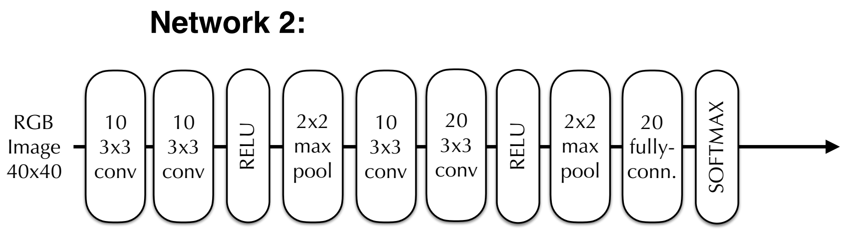
Hint: $\frac{1}{0.1} = 10$.

(c) Given code for Solver 1, describe how you can use this code to construct a solver with similar properties as Solver 2: Your solver should use $m = 4$ data points as input, return at most 1 model, require 1.01 time units to run, and evaluating one model on a single data point requires 0.01 time units.

[2] Out of the 4 solutions, at most one will be correct. The other solutions are introduced by the way the solver models the problem. However, you do not know which model is correct without evaluating it on all data points.

5 Deep Learning, 3 points





(a) Consider the two convolutional neural networks shown above. Which of the two networks has fewer parameters? Assume that the convolutional layers use padding and that no bias terms are used.

Explain your answer!

(b) Assume that you want to convert one of the two networks into a fully convolutional neural network. In order to do so, you will need to replace the fully connected layer with one convolutional layer. How many convolutional filters are needed and what is their size such that center pixel of a 40×40 input image produces the same result? In other words, you will need to apply N $k \times k$ filters. What are the values for N and k?

Explain your answer!

(c) In the context of generative modelling, draw the general architecture of a Variational Autoencoder (VAE). Explain the purpose of all its components. Which parts are only needed during testing?

6 Triangulation, 3 points

Consider two images with corresponding projection matrices $\mathtt{P}_1$ and $\mathtt{P}_2$. Using SIFT features, we have matched the 2D keypoint u_1 in image 1 with the 2D keypoint u_2 in image 2. We now want to use triangulation to compute the position of the corresponding 3D point X.

(a) Assuming that u_1 and u_2 perfectly correspond to the projection of X into image 1 and 2, respectively. Derive the geometric relations between u_1, u_2, X, $\mathtt{P}_1$, and $\mathtt{P}_2$ that need to hold.

Explain your answer and make sure that you explain all variables that you introduce.

(b) Starting from the relations in (a), derive a minimal linear solver for triangulation: Describe how to setup a system of linear equations $\mathtt{M}\theta = b$ such that solving the system for θ yields the 3D coordinates of the 3D point X.

Explain your answer and make sure that you explain all variables that you introduce. It is sufficient to explain how you obtain the linear system and what the unknown are that you need to solve for. You do not need to write down the actual equations.

(c) Consider the case of three images with corresponding projection matrices $\mathtt{P}_1$, $\mathtt{P}_2$, $\mathtt{P}_3$ and 2D point positions u_1, u_2, u_3. You know that each pair u_i and u_j ($i \neq j$) fulfils the epipolar constraint, i.e., u_j lies on the epipolar line of u_i in image j and vice versa.
RANSAC-based triangulation using the minimal solver from (b) returns exactly 2 inliers. Can you trust the solution?

Justify your answer.