Formulas IN4080 Fall 2022, relevant to lectures 1-7, 12-13

Jan Tore Lønning

I have got questions regarding which formulas you have to remember to the exam. As we see it, there aren't that many formulas. However, equally important as the formulas, are the interpretion of them and how they are used. There is not much use in rembering a formula if you don't understand what the letter stand for and how the formula should be applied. We give a list of formulas you are supposed to know, but you have to look at lectures or book to understand them.

Week 1

- Median
- Mean
- Variance
- Frequency
- Conditional frequency

Week2

Assumed backround in probabilities

- Estimating probability by frequency
- Conditional probability
- Independence
- Bayes' theorem
- Expectation
- Variance
- Standard deviation

Week 3

- Naïve Bayes
- Maximum likelihood estimation
- Laplace and Lidstone smoothing
- Accuracy
- Precision
- Recall
- F_1-score

Week 4

- Mean Square Error: $\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$
- Logistic formula (sigmoid): $y = \sigma(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{e^z+1}$
- Cross-entropy loss:

$$L_{CE}(\vec{w}) = -\log \prod_{i=1}^{m} P(y^{(i)}|\vec{x}^{(i)}) = \sum_{i=1}^{m} -\log P(y^{(i)}|\vec{x}^{(i)})$$

- Logistic regression, update: $\boldsymbol{w} \leftarrow (\boldsymbol{w} - \eta(\hat{y} - y)\boldsymbol{x})$

Week 5

- Multinomial Logistic Regression, lso called maximum entropy (maxent) classifier, or softmax regression: $P(C_j|\vec{x}) = \frac{e^{\overrightarrow{w_j}\cdot\vec{x}}}{\sum_{i=1}^{k} e^{\overrightarrow{w_i}\cdot\vec{x}}}$
- The accompanying update rule: $w_{i,j} = w_{i,j} - \eta(\hat{y}_i - y_i)x_j$

Week 6

- N-gram language model
- Perplexity score
- Interpolation
- Hidden Markov Model, bigram and n-gram
- Maximum Entropy Markov Model for tagging:
  - $\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} P(t_1^n|w_1^n) \approx \underset{t_1^n}{\operatorname{argmax}} \prod_{i=1}^{n} P(t_i|\, t_{i-k}^{i-1} w_{i-m}^{i+m})$
- Macro- and micro-average

Week 7

- Cosine similarity
- idf
- tf-idf
- analogical parallelograms

Week 12-13 (lecture 12)

- Formulas for simple feed-forward network, slide 12
- RelU
- For skip-gram w/ negative sampling: the learning objective, slide 43

Week 14 (lecture 13)

- Formulas for simple RNN, slide 6
- With softmax (slide 9, slide 12)
- (for encoder-decoder, slide 29 and attention, slide 32 and slide 33, do you not have to memorize formulas, but you should try to understand them and be able to answer questions about them.)