Grades

Print Print

Grade Item	Points	Weight Achieved	Grade	Comments and Assessments
Homework		50 / 50	100 %	
HW01	10 / 10	8.5 / 8.5	100 %	Overall Feedback Pros: • Comprehensive with good analysis and comparison Cons: • Plot multiple methods in 1 figure for good comparison. You shou even though it's optional. • It's nicer to recognize that Line 3 has a large error because of the test set is inconsistent with the training set) and not due to over • It would be nicer to have a plot of time/loss for the x/y axis for t speed vs. loss. (you need multiple methods on the same figure fo

				• It would be nicer to explain why the third method (analytic grad) method (finite difference) • It would be nicer to try and predict which method would scale b complicated model, high dimensional, etc. and explain. Overall Feedback Pros: • Comprehensive with good analysis and comparison Cons: • Plot multiple methods in 1 figure for good comparison. You shou even though it's optional. • It's nicer to recognize that Line 3 has a large error because of the test set is inconsistent with the training set) and not due to over • It would be nicer to have a plot of time/loss for the x/y axis for t speed vs. loss. (you need multiple methods on the same figure fo • It would be nicer to explain why the third method (analytic grad) method (finite difference) • It would be nicer to try and predict which method would scale b complicated model, high dimensional, etc. and explain.
HW02	10 / 10	8.5 / 8.5	100 %	Overall Feedback • 18 pages. Really? • (+0.5) Additional test for sigma=0.2 • I don't think that section 3.5 is very useful. • (+0.5) Extra experiments with different numbers of points in eac • (+1) Nice convergence plots in Fig 5, 6, 7, 8, 9, 10 • (-0.5) Missing the analysis why the Training error is lower than th • Fig 7a: the similarity score should go to 1 here. Double check?

				• The experiment in Fig 9 is not very different from Fig 8. • (+0.5) Nice PCA usage and plot (Fig 12). Overall Feedback • 18 pages. Really? • (+0.5) Additional test for sigma=0.2 • I don't think that section 3.5 is very useful. • (+0.5) Extra experiments with different numbers of points in eac • (+1) Nice convergence plots in Fig 5, 6, 7, 8, 9, 10 • (-0.5) Missing the analysis why the Training error is lower than th • Fig 7a: the similarity score should go to 1 here. Double check? • The experiment in Fig 9 is not very different from Fig 8. • (+0.5) Nice PCA usage and plot (Fig 12).
HW03	13 / 10	11.05 / 8.5	130 %	Overall Feedback • (+2) For PCA and TSNE figures and analysis • (+1) For the acc/loss changing with different bins value graph • A visualization of the Naive Bayes's classifier would be nice. Extr the decision boundary (on the PCA or TSNE figure). • You should clean up the later cells of the notebook. There is a (?) should clean this up. • Putting the important result (Table 1) as early as possible => Nic • Some keywords used in research based on your description: "Cla data imbalanced. "Feature Distribution Differences" => Distribut Dependence: It actually works surprisingly well even when this a can google some articles about this phenomenon. • No code for part 3 (posterior)? • (+3) Analysis for Part 3: correct. It would be better if you could r of having P(c | X), as it provides you with a full distribution with (variance). In contrast, modern NN can only provide a peak (a mo

				distribution) or a pseudo-distribution (the soft-max of NN does n Prof. Kobus's PGM class to see this :D (26/20) Overall Feedback • (+2) For PCA and TSNE figures and analysis • (+1) For the acc/loss changing with different bins value graph • A visualization of the Naive Bayes's classifier would be nice. Extr the decision boundary (on the PCA or TSNE figure). • You should clean up the later cells of the notebook. There is a (?) should clean this up. • Putting the important result (Table 1) as early as possible => Nic • Some keywords used in research based on your description: "Cla data imbalanced. "Feature Distribution Differences" => Distribut Dependence: It actually works surprisingly well even when this a can google some articles about this phenomenon. • No code for part 3 (posterior)? • (+3) Analysis for Part 3: correct. It would be better if you could r of having P(c | X), as it provides you with a full distribution with (variance). In contrast, modern NN can only provide a peak (a mo distribution) or a pseudo-distribution (the soft-max of NN does n Prof. Kobus's PGM class to see this :D (26/20)
HW04	8.5 / 10	7.23 / 8.5	85 %	Overall Feedback • The Github link was not found. • (-0.5) Fig 1, dataset 3: This is not overfitting since the model is n of the figures in Fig1). You can only conclude about overfitting w plateau and the test error is increasing.

				• (-0.5) Fig 1, dataset 4: this does not suggest overfitting. This ma or a test set that is different from the training set (distribution sh • The figure text is too small. Try increasing the font size a bit to a in your documents. At least for the headers and axis. Your figure zooming in, so I suggest changing the dpi of matplotlib to higher • (+2) Kudos for trying different values of lr and a number of epoc • (-3) Wrong conclusion about NN is worse than Naive Bayes. I su improper choice of learning rate. I gave you some extra details in • (+0.5) Fig 3 is of limited utility. Same with tables 3-6. Overall Feedback • The Github link was not found. • (-0.5) Fig 1, dataset 3: This is not overfitting since the model is n of the figures in Fig1). You can only conclude about overfitting w plateau and the test error is increasing. • (-0.5) Fig 1, dataset 4: this does not suggest overfitting. This ma or a test set that is different from the training set (distribution sh • The figure text is too small. Try increasing the font size a bit to a in your documents. At least for the headers and axis. Your figure zooming in, so I suggest changing the dpi of matplotlib to higher • (+2) Kudos for trying different values of lr and a number of epoc • (-3) Wrong conclusion about NN is worse than Naive Bayes. I su improper choice of learning rate. I gave you some extra details in • (+0.5) Fig 3 is of limited utility. Same with tables 3-6.
HW05	14 / 10	11.9 / 8.5	140 %	Overall Feedback • (-1) Fig 1 The model is not converged yet (need a flat tail on the • (+1) Fill_between • (+0.5) Additional experiments Table 2,3. Table 2 is a bit suspiciou observe a better result, even though the model hasn't converged

- (-1) You should choose a larger NN (so the medium has the best
  demonstrate that bigger is not always better. This caused Ming t
  conclusion in 2.3 (when the NN is too large, even adding regular
  enough to help. Refer to the deep models before ResNet).
- (+0.5) Nice analysis. Seulgi is correct, but you should focus more
  exp, the larger NN is better. To make any general conclusion like
  lot of data on multiple datasets, etc.
- There is no additional point for the dropout exp because the mo
  so dropout only hurts the performance. You should only add reg
  observe a large train-val gap, and you want to close the gap.
- (+2) Optional Q3
- (+2) Optional Q4. You should try to explain why ResNet is better
  of the Residual in the architecture design). And the fig definitely
  potentially.

Overall Feedback

- (-1) Fig 1 The model is not converged yet (need a flat tail on the
- (+1) Fill_between
- (+0.5) Additional experiments Table 2,3. Table 2 is a bit suspiciou
  observe a better result, even though the model hasn't converged
- (-1) You should choose a larger NN (so the medium has the best
  demonstrate that bigger is not always better. This caused Ming t
  conclusion in 2.3 (when the NN is too large, even adding regular
  enough to help. Refer to the deep models before ResNet).
- (+0.5) Nice analysis. Seulgi is correct, but you should focus more
  exp, the larger NN is better. To make any general conclusion like
  lot of data on multiple datasets, etc.
- There is no additional point for the dropout exp because the mo
  so dropout only hurts the performance. You should only add reg
  observe a large train-val gap, and you want to close the gap.
- (+2) Optional Q3
- (+2) Optional Q4. You should try to explain why ResNet is better
  of the Residual in the architecture design). And the fig definitely

				potentially.
HW06	12.5 / 10	9.38 / 7.5	125 %	Overall Feedback • Thank you for the theory and formula reminders. A bit unnecess • Fig 1,2: Train loss does not flatten => not converge yet. You corr there is no point loss. Fig 4 is still not converged, though. • (-0.5) You should plot Fig 1 and 2 in the same plot for compariso the instruction • (+2) Good analysis and explanation. Optional further reading: https://proceedings.neurips.cc/paper/2020/hash/288cd256795 Abstract.html • (+1) K-folds Overall Feedback • Thank you for the theory and formula reminders. A bit unnecess • Fig 1,2: Train loss does not flatten => not converge yet. You corr there is no point loss. Fig 4 is still not converged, though. • (-0.5) You should plot Fig 1 and 2 in the same plot for compariso the instruction • (+2) Good analysis and explanation. Optional further reading: https://proceedings.neurips.cc/paper/2020/hash/288cd256795 Abstract.html • (+1) K-folds
HW07 (Bonus)	3 / 10	2.25		Overall Feedback • No point was added for the paper discussion • (+3) missing extra credits for the Assignment 1 that was missing

				Overall Feedback • No point was added for the paper discussion • (+3) missing extra credits for the Assignment 1 that was missing
Quizzes		10 / 10	100 %	
Quiz 01	10 / 10	10 / 10	100 %	
Activities		9.7 / 10	97 %	
Activity 01	10 / 10	3 / 3	100 %	
Activity 02	10 / 10	1 / 1	100 %	
Activity 03	10 / 10	1 / 1	100 %	
Activity 04	10 / 10	1 / 1	100 %	
Activity 05	10 / 10	1 / 1	100 %	
Activity 06	9 / 10	2.7 / 3	90 %	
Midterms		9.28 / 10	92.8 %	

	Midterm 1 raw score	59 / 70	0 / 0	84.29 %
	Midterm 1	92.8 / 100	9.28 / 10	92.8 %
Final			17.98 / 20	89.9 %
	Final	89.9 / 100	17.98 / 20	89.9 %
	Final raw score	87.5 / 100	0 / 0	87.5 %
SCS bonus (Bonus)		10 / 10	2	