

HW2 Report

1.1 - Building a Linear Classifier

Perceptron Loss

Best dev accuracy: 64.72%

Best devtest accuracy: 57.51%

1.2 - Hinge Loss

Hinge Loss

Best dev accuracy: 67.27%

Best devtest accuracy: 58.98%

1.3 - Feature Weight Analysis

Rank	Label = 0	Label = 1	Label = 2
1	listless (1.69)	Cletis (0.68)	vividly (1.93)
2	lacks (1.62)	1950s (0.65)	pleasant (1.81)
3	tiresome (1.6)	Trooper (0.63)	thought-provoking (1.76)
4	hardly (1.55)	Warren (0.63)	wonderfully (1.74)
5	lacking (1.55)	Greek (0.58)	treat (1.63)
6	uneven (1.53)	Big (0.58)	delight (1.45)
7	pointless (1.5)	Island (0.56)	touching (1.44)
8	disingenuous (1.49)	Bartlett (0.56)	laughed (1.44)
9	lousy (1.49)	stake (0.56)	miracle (1.44)
10	sloppy (1.48)	Pete (0.56)	gorgeously (1.41)

Average feature weight: 0.0716

The features with the highest weights for the negative and positive labels make sense for the most part. The negative (label = 0) features with the ten highest weights are all adjectives and words one would commonly expect to find in critiques of films and their components (actors, plot, production). Similarly, the positive (label = 2) features with the ten highest weights are also all adjectives with positive connotations as one would expect. It's interesting to note that many of the neutral features (label = 1) are in fact proper nouns such as Greek, Bartlett, Pete, etc. This makes sense given that these nouns do not inherently have a positive or negative connotation.

1.4 - Error Analysis

(sentence / gold standard label / predicted label / error category)

- *To say this was done better in Wilder 's Some Like It Hot is like saying the sun rises in the east .* **0 2 - analogy**

- *A gripping , searing portrait of a lost soul trying to find her way through life .* **1 2 - wrong annotation** (should be 2)
- *Moretti 's compelling anatomy of grief and the difficult process of adapting to loss .* **0 2 - wrong annotation** (should be 2)
- *Scores no points for originality , wit , or intelligence .* **0 2 - negation**
- *Full of witless jokes , dealing in broad stereotypes and outrageously unbelievable scenarios , and saddled with a general air of misogyny* **1 0 wrong annotation** (should be 0)
- *It has all the excitement of eating oatmeal .* **0 2 - analogy**
- *It 's mighty tedious for the viewer who has to contend with unpleasant characters , hit-and-miss performances and awkwardly staged scenes .* **1 0 wrong annotation** (should be 0)
- *An occasionally funny , but overall limp , fish-out-of-water story .* **0 2 - conjunction error**
- *A sober and affecting chronicle of the leveling effect of loss .* **1 2 - wrong annotation** (should be 2)
- *More maudlin than sharp .* **1 0 - wrong annotation** (should be 0)
- *It 's sweet , harmless , dumb , occasionally funny and about as compelling as a fishing show .* **1 2 analogy**
- *I sympathize with the plight of these families , but the movie does n't do a very good job conveying the issue at hand .* **0 2 negation**
- *It 's inoffensive , cheerful , built to inspire the young people , set to an unending soundtrack of beach party pop numbers and aside from its remarkable camerawork and awesome scenery , it 's about as exciting as a sunburn .* **0 2 multiple sentiments**
- *Hilariously inept and ridiculous .* **2 0 - type of negation ("hilariously")/contains negative adjectives**
- *While the Resident Evil games may have set new standards for thrills , suspense , and gore for video games , the movie really only succeeds in the third of these .* **0 2 - multiple sentiments/conjunctive error**
- *As unseemly as its title suggests .* **2 0 - wrong annotation** (should be 0)
- *While it 's genuinely cool to hear characters talk about early rap records -LRB- Sugar Hill Gang , etc. -RRB- , the constant referencing of hip-hop arcana can alienate even the savviest audiences .* **0 2 multiple sentiments/conjunctive error**
- *Impostor has a handful of thrilling moments and a couple of good performances , but the movie does n't quite fly .* **0 2 - conjunction error**
- *A painfully funny ode to bad behavior .* **2 0 contains negative adjectives**
- *It 's somewhat clumsy and too lethargically paced -- but its story about a mysterious creature with psychic abilities offers a solid build-up , a terrific climax , and some nice chills along the way .* **0 2 wrong annotation/multiple sentiments** (should be 1 or 2)

1.5 - Feature Engineering

Template 1: Using a lemmatizer:

Used a lemmatizer on every word in the corpus and made the "lemmatized" form of the word a unigram binary feature, i.e "cars 2" becomes "car-2".

Best dev accuracy: 67.09%

Best devtest accuracy: 60.07% (+1.09%) from 1.2 result.

Learned Feature Weights:

Average feature weight: 0.0917

For each label I noticed that the features with the ten highest weights each had greater weights than those reported in 1.3. This makes sense given that the lemmatize process reduced the number of features. As such, many training examples that would have otherwise been represented as unique features were now represented by a single feature, i.e cars, car, car's, cars' => car. Moreover, given that we implemented hinge-loss, a lesser number of features meant that each feature had to compensate slightly more weight.

Template 2: Using a POS Tagger:

Only created unigram binary features for words that were classified as adjectives, adverbs, conjunctions, and nouns by a POS tagger, and ignored all others words in a given training example.

Best dev accuracy: 67.09%

Best devtest accuracy: 61.34% (+2.36%) from 1.2 result.

Learned Feature Weights:

Average feature weight: 0.0837

This feature template also led to a higher average feature weight than that which was produced in 1.3. It is likely that this increase can be attributed to the lesser number of features this template created, thus, as with template 1, features had to compensate with a slightly higher weight to account for hinge loss.