Class schemes

Scheme1: male vs. female vs. transmale vs. transfemale vs. other

Scheme2: cisgender vs. non-cisgender

Scheme3: male vs. female vs. non-cisgender

Dimensionality reduction

-Variance threshold: remove features that have same value for more than 80% of users

-Principle components: reduce to n\_features = n\_users

SVM variants

-SVC: one-vs-one

-LinearSVC: one-vs-rest

Other details

-Train/test split: 10% held out for test

-10-fold cross validation

-RBF, poly parameter settings: C = 10, gamma = 0.001

-Linear parameter settings: C = 0.01

-Sigmoid parameter settings: C = 1e04, gamma = 1e-06

**Scheme1**

*Zero-R*: 39.2%

*Mean CV accuracy*

|  |  |  |  |
| --- | --- | --- | --- |
|  | All features  (~65k features) | VarianceThreshold (2838 features) | PC  (3047 features) |
| SVC (kernel = rbf) |  | 72% | 72% |
| SVC (kernel = linear) |  | 68% |  |
| SVC (kernel = poly) |  | 69% |  |
| SVC (kernel = sigmoid) |  | 68% |  |
| LinearSVC |  | 64% | 71% |
|  |  |  |  |

*Test accuracy*

|  |  |  |  |
| --- | --- | --- | --- |
|  | All features  (~65k features) | VarianceThreshold (2838 features) | PC  (3047 features) |
| SVC (kernel = rbf) |  | 72.1% (220/305) | 71.8% (218/205) |
| SVC (kernel = linear) |  | 40% (122/305) |  |
| SVC (kernel = poly) |  | 55.4% (169/305) |  |
| SVC (kernel = sigmoid) |  | 73% (223/305) |  |
| LinearSVC |  | 67.5% (206/305) | 70.5% (215/305) |

**Scheme2**

*Zero-R*: 55%

*Mean CV accuracy*

|  |  |  |  |
| --- | --- | --- | --- |
|  | All features  (~65k features) | VarianceThreshold (2838 features) | PC  (3047 features) |
| SVC |  | 88% | 81% |
| LinearSVC |  | 86% | 89% |

*Test accuracy*

|  |  |  |  |
| --- | --- | --- | --- |
|  | All features  (~65k features) | VarianceThreshold (2838 features) | PC  (3047 features) |
| SVC |  | 89.2% (272/305) | 89.8% (274/305) |
| LinearSVC |  | 86.9% (265/305) | 89.8% (274/305) |

**Scheme3**

*Zero-R*: 45%

*Mean CV accuracy*

|  |  |  |  |
| --- | --- | --- | --- |
|  | All features  (~65k features) | VarianceThreshold (2838 features) | PC  (3047 features) |
| SVC |  | 77% | 67% |
| LinearSVC |  | 73% | 78% |

*Test accuracy*

|  |  |  |  |
| --- | --- | --- | --- |
|  | All features  (~65k features) | VarianceThreshold (2838 features) | PC  (3047 features) |
| SVC |  | 76.1% (232/305) | 76.1% (232/305) |
| LinearSVC |  | 72.5% (221/305) | 75.7% (231/305) |