Modern Computational Statistics, Problem Set 1

September 23, 2019 Due 10/07/2019

Problem 1.

(1) Show that $X \sim \mathcal{N}(0, 1)$ is the maximum entropy distribution such that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$.

(2) Generalize the result in (1) for the maximum entropy distribution given the first k moments, *i.e.*, $\mathbb{E}X^i = m_i,\ i = 1, \ldots, k$.

Problem 2.

Use the following code to generate covariate matrices X

```
import numpy as np
np.random.seed(1234)

n = 100
X = np.random.normal(size=(n,2))
```

(1) Generate $n = 100$ observations Y following the logistic regression model with true parameter $\beta_0 = (-2, 1)$.

(2) Find the MLE using the iteratively reweighted least square algorithm.

(3) Repeat (1) and (2) for 100 instances. Compare the MLEs with the asymptotical distribution $\hat{\beta} \sim \mathcal{N}(\beta_0, \mathcal{I}^{-1}(\beta_0))$. Present your result with a scatter plot for MLEs with contours for the pdf of the asymptotical distribution.

(4) Try the same for $n = 10000$. Does the asymptotical distribution provide a better fit to the MLEs? You can use the empirical covariance matrix of the MLEs for comparison.

Problem 3.

Similarly as in Problem 2, generate a large covariate matrix X with 100000 instances and 100 features, and response Y with true parameter β_0

```
import numpy as np
np.random.seed(1234)

n, d = 100000, 100
X = np.random.normal(size=(n,d))
beta_0 = np.random.normal(size=d)
```

(1) Compare gradient descent and nesterov's accelerated gradient descent.

(2) Compare vanilla stochastic gradient descent with different adaptive stochastic gradient descent methods, including AdaGrad, RMSprop, and Adam. Using minibatch sizes $32, 64, 128$.

(3) Bonus question. Generate a random mask matrix M as follows and use it to sparsify

the covariance matrix X

```
np.random.seed(1234)

sparse_rate = 0.3
M = np.random.uniform(size=(n,d)) < sparse_rate
X[M] = 0.
```

Repeat your experiments in (2), and compare with the results for the full covariance matrix.